

THE ORIGIN OF PLANT CHEMODIVERSITY - CONCEPTUAL AND EMPIRICAL INSIGHTS

EDITED BY: Kazuki Saito

PUBLISHED IN: Frontiers in Plant Science





frontiers

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714

ISBN 978-2-88963-923-6

DOI 10.3389/978-2-88963-923-6

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

THE ORIGIN OF PLANT CHEMODIVERSITY - CONCEPTUAL AND EMPIRICAL INSIGHTS

Topic Editor:

Kazuki Saito, RIKEN Center for Sustainable Resource Science (CSRS), Japan

Citation: Saito, K., ed. (2020). The Origin of Plant Chemodiversity - Conceptual and Empirical Insights. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88963-923-6

Table of Contents

- 05 Editorial: The Origin of Plant Chemodiversity – Conceptual and Empirical Insights**
Kazuki Saito
- 07 Induced, Imprinted, and Primed Responses to Changing Environments: Does Metabolism Store and Process Information?**
Jens Schwachtje, Sarah J. Whitcomb, Alexandre Augusto Pereira Firmino, Ellen Zuther, Dirk K. Hinch and Joachim Kopka
- 20 The Recruitment Model of Metabolic Evolution: Jasmonate-Responsive Transcription Factors and a Conceptual Model for the Evolution of Metabolic Pathways**
Tsubasa Shoji
- 32 Chemodiversity of the Glucosinolate-Myrosinase System at the Single Cell Type Resolution**
Shweta Chhajed, Biswapriya B. Misra, Nathalia Tello and Sixue Chen
- 43 Evolutionary Developments in Plant Specialized Metabolism, Exemplified by Two Transferase Families**
Hiroaki Kusano, Hao Li, Hiroshi Minami, Yoshihiro Kato, Homare Tabata and Kazufumi Yazaki
- 49 Acceleration of Mechanistic Investigation of Plant Secondary Metabolism Based on Computational Chemistry**
Hajime Sato, Kazuki Saito and Mami Yamazaki
- 65 Large-Scale Profiling of Saponins in Different Ecotypes of *Medicago truncatula***
Zhentian Lei, Bonnie S. Watson, David Huhman, Dong Sik Yang and Lloyd W. Sumner
- 77 Assessing Chemical Diversity in *Psilotum nudum* (L.) Beauv., a Pantropical Whisk Fern That Has Lost Many of its Fern-Like Characters**
Dunja Šamec, Verena Pierz, Narayanan Srividya, Matthias Wüst and B. Markus Lange
- 95 Formation of Flavonoid Metabolons: Functional Significance of Protein-Protein Interactions and Impact on Flavonoid Chemodiversity**
Toru Nakayama, Seiji Takahashi and Toshiyuki Waki
- 107 Evolutionary Diversification of Primary Metabolism and its Contribution to Plant Chemical Diversity**
Hiroshi A. Maeda
- 115 Identification and Functional Characterization of Genes Involved in the Biosynthesis of Caffeoylquinic Acids in Sunflower (*Helianthus annuus* L.)**
Ketthida Cheeverunnapakul, Gholamreza Khaksar, Pawinee Panpetch, Patwira Boonjing and Supaart Sirikantaramas
- 129 The Origin and Evolution of Plant Flavonoid Metabolism**
Keiko Yonekura-Sakakibara, Yasuhiro Higashi and Ryo Nakabayashi
- 145 Atypical Myrosinase as a Mediator of Glucosinolate Functions in Plants**
Ryosuke Sugiyama and Masami Y. Hirai

- 159 ***Molecular Origins of Functional Diversity in Benzyloquinoline Alkaloid Methyltransferases***
Jeremy S. Morris and Peter J. Facchini
- 181 ***Corrigendum: Molecular Origins of Functional Diversity in Benzyloquinoline Alkaloid Methyltransferases***
Jeremy S. Morris and Peter J. Facchini
- 182 ***Biosynthesis and Emission of Stress-Induced Volatile Terpenes in Roots and Leaves of Switchgrass (Panicum virgatum L.)***
Andrew Muchlinski, Xinlu Chen, John T. Lovell, Tobias G. Köllner, Kyle A. Pelot, Philipp Zerbe, Meredith Ruggiero, LeMar Callaway III, Suzanne Laliberte, Feng Chen and Dorothea Tholl
- 198 ***Spatio-Temporal Metabolite and Elemental Profiling of Salt Stressed Barley Seeds During Initial Stages of Germination by MALDI-MSI and μ -XRF Spectrometry***
Sneha Gupta, Thusitha Rupasinghe, Damien L. Callahan, Siria H. A. Natera, Penelope M. C. Smith, Camilla B. Hill, Ute Roessner and Berin A. Boughton
- 215 ***Terpene Synthases as Metabolic Gatekeepers in the Evolution of Plant Terpenoid Chemical Diversity***
Prema S. Karunanithi and Philipp Zerbe
- 238 ***Taxonomically Informed Scoring Enhances Confidence in Natural Products Annotation***
Adriano Rutz, Miwa Dounoue-Kubo, Simon Ollivier, Jonathan Bisson, Mohsen Bagheri, Tongchai Saesong, Samad Nejad Ebrahimi, Kornkanok Ingkaninan, Jean-Luc Wolfender and Pierre-Marie Allard
- 253 ***Untargeted Metabolomics of Nicotiana tabacum Grown in United States and India Characterizes the Association of Plant Metabolomes With Natural Climate and Geography***
Dong-Ming Ma, Saiprasad V. S. Gandra, Raman Manoharlal, Christophe La Hovary and De-Yu Xie
- 271 ***Contribution of Functional Divergence Through Copy Number Variations to the Inter-Species and Intra-Species Diversity in Specialized Metabolites***
Kazumasa Shirai and Kousuke Hanada
- 277 ***Molecular Basis of C-30 Product Regioselectivity of Legume Oxidases Involved in High-Value Triterpenoid Biosynthesis***
Much Zaenal Fanani, Ery Odette Fukushima, Satoru Sawai, Jianwei Tang, Masato Ishimori, Hiroshi Sudo, Kiyoshi Ohyama, Hikaru Seki, Kazuki Saito and Toshiya Muranaka
- 293 ***Evolution of Structural Diversity of Triterpenoids***
Pablo D. Cárdenas, Aldo Almeida and Søren Bak
- 303 ***The Evolution of Flavonoid Biosynthesis: A Bryophyte Perspective***
Kevin M. Davies, Rubina Jibran, Yanfei Zhou, Nick W. Albert, David A. Brummell, Brian R. Jordan, John L. Bowman and Kathy E. Schwinn



Editorial: The Origin of Plant Chemodiversity – Conceptual and Empirical Insights

Kazuki Saito^{1,2*}

¹ RIKEN Center for Sustainable Resource Science, Yokohama, Japan, ² Plant Molecular Science Center, Chiba University, Chiba, Japan

Keywords: phytochemicals, plant metabolism, metabolomics, biosynthesis, specialized metabolite, evolution, chemodiversity, plant chemicals

Editorial on the Research Topic

The Origin of Plant Chemodiversity – Conceptual and Empirical Insights

Whenever one looks plant-made chemicals, everyone is fascinated with their huge chemodiversity, being estimated at over one million metabolites (Afendi et al., 2012), and then impressed how people lives rely on those plant chemicals as food, drugs, flavors, cosmetics, and industrial raw materials. Although it is generally accepted that such vast plant chemodiversity is the consequence of the evolutionary history of plants as sessile organisms, the queries regarding the origin of plant chemodiversity are not fully addressed.

By taking the advantage of recently advanced theory and technology of genomics and associated fields (Rai et al., 2017), such as comparative genomics, metabolomics, bioinformatics and molecular evolution, one may address the questions mentioned above quite directly. Besides answering such fundamental questions, the knowledge gained will contribute to the development of sustainable society, which is formulated as 2030 Agenda for Sustainable Development and its 17 Sustainable Development Goals (United Nations, 2018).

Only outstanding experts in this field were invited to contribute their articles to this Research Topic. The collected papers widely cover the topic from conceptual discernment to empirical insights. Finally, in total, 23 articles were published and divided into several categories: 9 in Review, 2 in Mini Review, 2 in Perspective, 1 in Hypothesis and Theory, 8 in Original Research, and 1 in Correction.

Metabolomics has now maturely held an indispensable position to decipher the chemodiversity of plants. However, peak annotation of acquired data by untargeted analysis is still challenging. Rutz et al. report an improved method for annotation of natural products with taxonomic information. Spatio-temporal metabolite and elemental profiling were applied to investigate salt-stress response in barley (Gupta et al.). The impacts of natural climate and geography on the metabolome of tobacco is reported by Ma et al.. The paper by Šamec et al. deals with the metabolome of the Psilotales, which exhibit unique anatomical character in the fern lineage.

Several articles deal with an evolutionary consideration in the expansion of plant chemodiversity. Two Perspective articles discuss on evolution aspects: The report by Kusano et al. provides an evolutionary perspective regarding two transferases often involved in plant secondary (recently often referred to as “specialized”) metabolism; another perspective paper by Shirai and Hanada discusses the effects of copy number variations on functional divergence in the cross-species and inside-species diversity. Shoji proposes the evolutionary model based on recruitment of genes by focusing on jasmonate-responsive transcription factors. The Mini-Review article by Maeda provides insight on the evolutionary diversification of plant primary metabolism

OPEN ACCESS

Edited and reviewed by:

Kirsi-Marja Oksman-Caldentey,
VTT Technical Research Centre of
Finland Ltd, Finland

*Correspondence:

Kazuki Saito
kazuki.saito@riken.jp

Specialty section:

This article was submitted to
Plant Metabolism and Chemodiversity,
a section of the journal
Frontiers in Plant Science

Received: 05 May 2020

Accepted: 29 May 2020

Published: 30 June 2020

Citation:

Saito K (2020) Editorial: The Origin of
Plant Chemodiversity – Conceptual
and Empirical Insights.
Front. Plant Sci. 11:890.
doi: 10.3389/fpls.2020.00890

concerning the robustness of plant metabolism and further diversification in specialized metabolism.

Two papers deal with the novel concept and approach to extend our understanding of plant metabolism. The Hypothesis and Theory article by Schwachtje et al. discusses the imprinting and priming plant metabolism against upcoming environmental challenges as a novel concept for adaptation and diversification of plant metabolism responding to stresses. To obtain further insights on mechanisms of plant natural products biosynthesis, computational chemistry may play a central role—Sato et al. present an overview of cutting-edge aspects of application of computational chemistry on the biosynthetic mechanisms of plant specialized metabolites.

The rest of the 13 articles focuses on relatively the pathway-specific metabolism or metabolites with some generalization across the biosynthetic pathways. Four papers deal with flavonoids and phenylpropanoids. Yonekura-Sakakibara et al. discuss the origin of flavonoid biosynthetic enzymes with functional genomics implication in their comprehensive Review Article. Another Review Article (Davies et al.) provides the current information about the flavonoid pathway in the bryophytes (liverworts, hornworts, and mosses). The role of the flavonoid metabolites formed by specific protein-protein interactions of the biosynthetic enzymes is discussed in the Review Article (Nakayama et al.). The Original Research Article by Cheevarungrapakul et al. reports the genes for the biosynthesis of caffeoylquinic acids in sunflower.

Terpenoids form the largest sub-family of specialized metabolites with their immense possibility in the expansion of structural diversity. The Review Article by Karunanithi and Zerbe provides up-to-date knowledge on the functional diversity and molecular evolution of the plant terpene synthases widely across the plant kingdom. The Original Article contributed by Muchlinski et al. reports biosynthesis and emission of mixtures of monoterpenes and sesquiterpenes triggered by the generalist herbivore in Switchgrass. Three papers focus on triterpenoid saponins. Lei et al. studied the large-scale profiling of saponins in 201 ecotypes of *Medicago truncatula*. The Original Article by Fanani et al. reports the detailed investigation on the molecular basis for C-30 oxidation of triterpenoids leading to the production of high-value saponins such as glycyrrhizin in Legume licorice. Cárdenas et al. in their Mini-Review Article

discuss convergent and divergent evolution in triterpenoid biosynthesis, and the mechanisms increasing structural diversity within and across plant species.

Three contributions dealing with nitrogen- and sulfur-containing specialized metabolites are included. Morris and Facchini(a); Morris and Facchini(b) provide an overview of the functional diversification of methyltransferases in the biosynthesis of benzyloisoquinoline alkaloids. The Review by Chhajed et al. discusses the cellular and subcellular organization of the glucosinolates -myrosinase system, its chemodiversity and functions in different cell types, emphasizing single-cell-type studies. Sugiyama and Hirai provide an up-to-date information on the diversity and the role of atypical myrosinases beyond the classical model of glucosinolates -myrosinase system.

Overall, this Research Topic becomes an excellent anthology to exhibit state-of-the-art on the theme of Topic by contributions from world-experts of this field. The Research Topic can play a role as a sort of flagship topic of the Specialty Section of Plant Metabolism and Chemodiversity. The curiosity-driven fundamental research on the origin and evolution of diversification of plant chemicals is primarily essential. Besides, the basic knowledge obtained could contribute to solving the current global problems such as climate crisis and pandemic diseases. I hope this Research Topic could be a landmark of future research in this field.

AUTHOR CONTRIBUTIONS

KS contributes entirely to the paper.

FUNDING

This work was supported by the JSPS KAKENHI program (grant number 19H05652 to KS) and by the Strategic Priority Research Promotion Program of Chiba University.

ACKNOWLEDGMENTS

I thank all authors and reviewers who contributed to the success of this Research Topic. I also acknowledge the Frontiers Editorial Office for their technical support.

REFERENCES

- Afendi, F. M., Okada, T., Yamazaki, M., Hirai-Morita, A., Nakamura, Y., Nakamura, K., et al. (2012). KNApSACk family databases: integrated metabolite-plant species databases for multifaceted plant research. *Plant Cell Physiol.* 53:e1. doi: 10.1093/pcp/pcr165
- Rai, A., Saito, K., and Yamazaki, M. (2017). Integrated omics analysis of specialized metabolism in medicinal plants. *Plant J.* 90, 764–787. doi: 10.1111/tpj.13485
- United Nations (2018). *Sustainable Development Goals, Knowledge Platform*. Available online at: <https://sustainabledevelopment.un.org/> (accessed May 24, 2020).

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Saito. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Induced, Imprinted, and Primed Responses to Changing Environments: Does Metabolism Store and Process Information?

Jens Schwachtje*, Sarah J. Whitcomb, Alexandre Augusto Pereira Firmino, Ellen Zuther, Dirk K. Hincha and Joachim Kopka

Department of Molecular Physiology, Applied Metabolome Analysis, Max-Planck-Institute of Molecular Plant Physiology, Potsdam, Germany

OPEN ACCESS

Edited by:

Kazuki Saito,
RIKEN Center for Sustainable
Resource Science (CSRS), Japan

Reviewed by:

Alex Williams,
University of Sheffield,
United Kingdom
Masami Yokota Hirai,
RIKEN Center for Sustainable
Resource Science (CSRS), Japan

*Correspondence:

Jens Schwachtje
schw8je@gmail.com

Specialty section:

This article was submitted to
Plant Metabolism and
Chemodiversity,
a section of the journal
Frontiers in Plant Science

Received: 28 September 2018

Accepted: 23 January 2019

Published: 13 February 2019

Citation:

Schwachtje J, Whitcomb SJ,
Firmino AAP, Zuther E, Hincha DK
and Kopka J (2019) Induced,
Imprinted, and Primed Responses
to Changing Environments:
Does Metabolism Store and
Process Information?
Front. Plant Sci. 10:106.
doi: 10.3389/fpls.2019.00106

Metabolism is the system layer that determines growth by the rate of matter uptake and conversion into biomass. The scaffold of enzymatic reaction rates drives the metabolic network in a given physico-chemical environment. In response to the diverse environmental stresses, plants have evolved the capability of integrating macro- and micro-environmental events to be prepared, i.e., to be primed for upcoming environmental challenges. The hierarchical view on stress signaling, where metabolites are seen as final downstream products, has recently been complemented by findings that metabolites themselves function as stress signals. We present a systematic concept of metabolic responses that are induced by environmental stresses and persist in the plant system. Such metabolic imprints may prime metabolic responses of plants for subsequent environmental stresses. We describe response types with examples of biotic and abiotic environmental stresses and suggest that plants use metabolic imprints, the metabolic changes that last beyond recovery from stress events, and priming, the imprints that function to prepare for upcoming stresses, to integrate diverse environmental stress histories. As a consequence, even genetically identical plants should be studied and understood as phenotypically plastic organisms that continuously adjust their metabolic state in response to their individually experienced local environment. To explore the occurrence and to unravel functions of metabolic imprints, we encourage researchers to extend stress studies by including detailed metabolic and stress response monitoring into extended recovery phases.

Keywords: priming, stress response, stress signaling, metabolism, metabolic imprint, plant physiology

INTRODUCTION

Sessile plants are forced to respond to adverse biotic and abiotic conditions in their local environment. Depending on the nature and intensity of such conditions, a plant's physiology can change markedly, generally because of stress-activated signaling that leads to specific physiological responses. These responses protect against or mitigate deleterious effects of stress. In a top-to-bottom view, stress-related cues induce signaling cascades, followed by activities at genetic and protein levels. Metabolite changes are generally considered to be the last step in this event cascade. Typically, large parts of metabolism

are affected. Stress-related cues may directly influence enzyme activities that modify the metabolic state, independent of the transcription/translation machinery. Finally, metabolic changes result from cues that do not induce classical stress-signaling. Such direct changes of metabolism are caused by the fluctuating physico-chemical environment of a plant, such as varying climate parameters, e.g., temperature, or soil properties (Hawkes and Sullivan, 2001; De Deyn et al., 2008; Sampaio et al., 2015; Onwuka and Mang, 2018).

Most environmental stresses are transient, such as temporally limited temperature extremes or drought phases, insect attacks or microbial infections. After a stress has ended, plants may recover from the stress and reset metabolism to growth and reproduction modes; however, recovery may not be complete. Even short-term environmental stress may have long-lasting effects on the plant system. A growing number of studies suggest that storing information about a past stress event benefits plants by preparing them for the same or a similar stress in the future. This phenomenon is called priming (Frost et al., 2008; Gamir et al., 2014a; Conrath et al., 2015; Hilker et al., 2015; Mauch-Mani et al., 2017). Several plant priming mechanisms are well investigated (**Table 1**).

Priming mechanisms are described at different levels: at the epigenetic level (e.g., by DNA and histone modification) and at the transcript or protein level (e.g., persisting changes in the abundance of transcripts, including transcription factors, and proteins or modulation of enzyme activities). However, the metabolic level as a mediator of priming has remained largely unexplored, even though large parts of metabolism are altered during stress (e.g., Schwachtje and Baldwin, 2006; Bolton, 2009; Krasensky and Jonak, 2012; Fraire-Velázquez and Balderas-Hernández, 2013). Here, we hypothesize that persistent stress-induced changes in metabolite concentrations,

metabolite ratios, or metabolic fluxes represent a **metabolic imprint** of prior environmental impacts and that these imprints can prime responses to future environmental events. We present evidence that supports our hypothesis and suggest environmental shift experiments that not only monitor metabolic responses during a first stress exposure (the priming event) or during a second stress response (the primed response), but also monitor short- and long-term recovery phases after stress events. Such experimental designs may characterize and identify functions of **metabolic imprints** at the level of **metabolic priming**.

Metabolite-primed responses are only properly defined by the timing, nature, and dose of the preceding environmental change, the duration of the recovery period, and in addition by the nature and dose of the subsequent stress for which metabolism is primed. In agreement with generalized concepts of non-metabolic primed stress responses (Hilker et al., 2015), several scenarios of metabolite-primed stress responses are conceivable:

1. The primed response may be stronger than the non-primed response level. As a result, defense or tolerance mechanisms can be amplified.
2. The primed response rate may be accelerated and reach effective response levels earlier.
3. The primed response is initiated earlier. In this case, the system's response rate may remain unchanged, but the effective levels of response are reached earlier.
4. Primed responses may be triggered by a lower stress dose.

These scenarios are thought to be generally applicable and have been recently discussed (Hilker et al., 2015). In the following, we focus on the roles of metabolites during stress responses, recovery, and priming. We shortly highlight effects of metabolites at all system levels of plant physiology and subsequently review metabolic changes that are caused by stress and last during stress recovery as metabolic imprints. We link metabolic imprints to a wide range of abiotic and biotic stresses. Finally, we discuss experimental approaches that enable discovery of metabolic imprints and functional analyses of these imprints.

INDUCED METABOLIC RESPONSES

Changes in the biotic and abiotic environment are reflected by the metabolic state of a plant. Plants have a multitude of plastic responses hardwired into their genomes (Sultan, 2000). These responses are induced concomitantly with the stress and function as defense, tolerance, or repair mechanisms (e.g., Dangel and Jones, 2001; Suzuki et al., 2011; Schuman and Baldwin, 2015). These mechanisms can be defined as **stress-signaling dependent metabolic responses**. Additionally, physical and chemical conditions such as temperature and soil nutrients influence metabolism, albeit mostly to a lesser degree than stresses. Temperature affects all reaction and transport rates. Soil nutrients influence physiology according to their availability. Consequently, changes in the physico-chemical environment of a plant will cause concomitant metabolic responses which

TABLE 1 | Examples of stress priming scenarios. Many abiotic and biotic stresses lead to imprints that improve the plant's response to a subsequent stress.

Priming stress	Induction by	Reference
Biotic/insect	Feeding	Frost et al., 2008
	Oviposition	Bandoly et al., 2016
	Volatiles	Engelberth et al., 2004
Biotic/microorganism	Pathogens (SAR)	Jaskiewicz et al., 2011; Dempsey and Klessig, 2012; Zeier, 2013
	Rhizobacteria (ISR)	van Loon, 2007; van der Ent et al., 2009
	Symbiotic fungi	Pozo et al., 2009
Abiotic	Cold/freezing	Thomashow, 1999; Guy, 2003; Hinch and Zuther, 2014; Zuther et al., 2018
	Salt	Sani et al., 2013
	Heat	Stief et al., 2014; Bäurle, 2016
	Drought	Ding et al., 2012
Other stimuli	β -Aminobutyric acid Salicylic acid Seed priming with different techniques	Zimmerli et al., 2001 Jakab et al., 2001 Jisha et al., 2013; Donohue, 2009; Rasmann et al., 2011

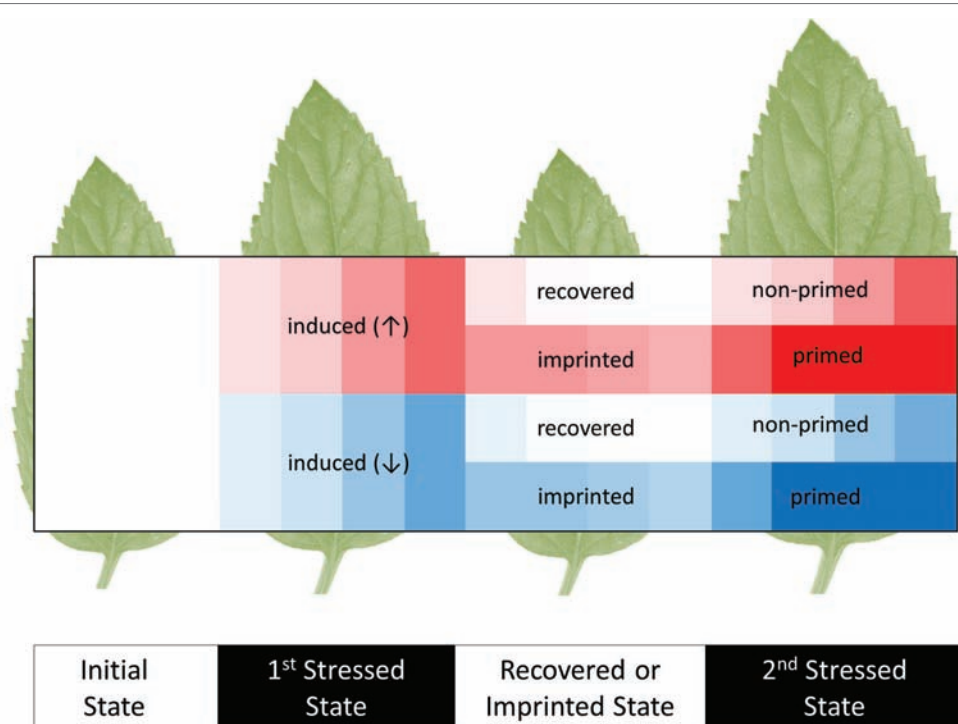


FIGURE 1 | Scheme of initial, induced, imprinted, and primed metabolic responses to recurrent abiotic or biotic stresses. The heat map shows increases (↑, red) and decreases (↓, blue) of metabolite levels relative to the pre-stressed, initial metabolic state (white). Most induced responses that are caused by a first stress period subsequently recover to the initial metabolic states. Few metabolic changes are retained. We argue that these metabolic imprints may be used as a memory that stores and processes information of preceding stress histories. These processes may contribute to stress priming. A primed response to a second stress enables organisms to cope better with recurrent stresses. Parts of the primed metabolic responses can be more sensitive, earlier, faster, or result in more extreme changes of metabolite levels. Note that imprinted metabolites may not necessarily have primed responses. Indeed, metabolic imprints during recovery without prior induction and primed responses without prior metabolic induction or imprinting are conceivable.

influence single or multiple nutrient fluxes through the plant system. These changes may ultimately become apparent as changes in metabolic pool sizes or fluxes. Therefore, **induced metabolic responses** need to be interpreted as the synergistic effects of stress-signaling dependent plant responses and of external physico-chemical influences.

The study of induced metabolic responses implies an *initial metabolic state* that transitions into an *induced metabolic state* (Figure 1); however, these states cannot be viewed as steady states, because they are integrated into non-static physiological processes governed by diurnal environmental cycles, circadian rhythms, and developmental progression of specific tissues and of the whole plant system. Due to these interactions, any observed induced metabolic response may represent a direct modification of metabolism that is on top of the underlying physiological programs of the plant.

METABOLIC IMPRINTS AND METABOLIC MEMORY

After an environmental stress has ceased, plant metabolism typically returns to a recovered state that is highly similar to the initial state (e.g., Hemme et al., 2014; Crisp et al., 2016; Pagter et al., 2017). A very basic metabolic perturbation and

recovery process may resemble a hysteresis curve, where the metabolic transition during the perturbation and recovery take characteristically different paths. The kinetics of metabolic remodeling are largely dependent on the nature of the environmental stress, transcriptional activities, the architecture of the affected metabolic networks, and the set of transport rates and enzyme activities that act on the induced metabolite levels. As an example of differential metabolic remodeling kinetics, glycolysis intermediates reached pre-stress levels quicker than TCA cycle intermediates in *Arabidopsis* roots upon recovery from oxidative stress (Lehmann et al., 2012).

Some induced metabolic responses may persist after the global metabolic state of the plant has recovered to the initial state (Figure 1). A simple case would be delayed adjustment of a metabolite to the initial state. Such a delay causes a metabolite to be more abundant at the onset of a second stress event. Furthermore, some induced metabolic responses may be effectively permanent or even cumulative, e.g., due to the absence of a catabolic pathway or sequestration mechanism (Mackie et al., 2013). Two phases of metabolic response are evident: (1) metabolic changes that are induced as immediate, specific responses to the stress, e.g., sugar levels, sucrose/hexose ratio, precursors for secondary metabolites or energy-related metabolites. During stress recovery (2) some of these changes may last as mid- or long-term imprints. Imprints may influence upcoming stress

responses. Here, we define *metabolic imprints* to encompass all metabolic changes that persist after recovery and thus differentiate the imprinted from the initial, pre-stress state.

According to this definition, metabolic imprints are indicators of past environmental conditions and/or stress events. In this sense, metabolic imprints may store information. Imprinted information becomes *metabolic memory*, when it is maintained and used by the plant system to improve future stress responses, for example to enhance or accelerate metabolite-induced signaling (Thellier and Lüttge, 2013). Metabolic imprints may be caused by one or more environmental events in the individual history of a plant. Metabolic imprints have been postulated previously to act as a stress memory (Bruce et al., 2007). A metabolic memory may act alone or more likely in synergy with priming and memory mechanisms at other system levels that are highlighted in the following.

METABOLITES INFLUENCE BIOLOGICAL SYSTEMS AT ALL LEVELS

If metabolites are involved in stress responses and represent stored information, metabolites should in turn influence metabolic or signaling pathways and other parts of plant physiology to modify a stress response. Generally, stress metabolism is seen as hierarchically organized, where external cues initiate signaling pathways that *via* transcription, translation, posttranslational modifications such as phosphorylation and further regulatory steps ultimately affect metabolism. Metabolites are generally thought to represent the downstream “end” products of this hierarchy. Interestingly, this view is currently complemented by findings that suggest bottom-to-top signaling mechanisms. Specific metabolites can exert regulatory influence or feedback on the stress-signaling network and physiology. Such mechanisms open possibilities for cross talk between stress-induced metabolites and other levels of physiological regulation (Bonawitz et al., 2012; Farre and Weise, 2012; Xiao et al., 2012; Gaudinier et al., 2015; Katz et al., 2015; Francisco et al., 2016; Malinovski et al., 2017). Further complexity is indicated by fluxes of central metabolism that are not necessarily explained by transcript abundances of the corresponding enzymes (Chubukov et al., 2013; Schwender et al., 2014). In addition to classical allosteric feedback responses, such as the suppression of enzyme activity by high levels of reaction products, metabolite ratios, and possibly also metabolite fluxes may thus play important roles by directly affecting multiple levels of the stress response hierarchy.

For example, branched chain amino acids are required for phosphorylation of G proteins during osmotic stress signaling in yeast (Shellhammer et al., 2017). Other findings suggest that certain primary metabolites can influence physiology at the transcriptional level. In yeast, Pinson et al. (2009) found that a metabolic intermediate of purine metabolism influences the interaction of transcription factors and thereby modulates purine and phosphate metabolism. Amino acids and polyamines are suggested to directly modify translation because they can

be sensed by translating ribosomes *via* interactions with nascent polypeptides, specifically with so-called arrest peptides (Seip and Innis, 2016). Furthermore, several metabolites are cofactors or co-substrates of chromatin-modifying enzymes and thus represent a potential regulatory interface between the metabolic and chromatin states of the cell (Shen et al., 2016; Van der Knaap and Verrijzer, 2016). For some primary metabolites, a role in chromatin modulation is suggested, e.g., fumarate, succinate, α -ketoglutarate, and acetyl-CoA. Fumarate, for example, is a competitive inhibitor of α -ketoglutarate, which is a co-substrate of histone demethylases and TET DNA methylases. Changes in cellular fumarate levels or ratios of fumarate, e.g., to ketoglutarate, may therefore contribute to altered histone modification. Different methylated histone residues are sensitive to changes in the α -ketoglutarate/succinate ratio (Van der Knaap and Verrijzer, 2016). These effects may be specific for certain genetic regions.

The TOR and SnRK kinases are sensors of the cellular energy state and can regulate large parts of metabolism. Plants adapt to changes in energy requirements during stress using these sensors (Baena-González, 2010; Hey et al., 2010; Rexin et al., 2015). The kinases are suggested to respond to sugars and other metabolites, even though the molecular mechanisms are not yet unraveled in all details. TOR is known to be regulated by nutrient sensing of nitrogen and carbon metabolites in plants, yeast, and mammals (e.g., Dobrenel et al., 2016; González and Hall, 2017). But plant defense metabolites of the glucosinolate family, 3-hydroxypropylglucosinolate, and/or its derivatives, can also activate TOR kinases (Malinovski et al., 2017). A precursor of plastidial isoprenoids that is induced by abiotic stress can induce nuclear stress-responsive genes *via* retrograde signaling (Xiao et al., 2012). The mediator complex, which regulates gene transcription, is involved in phenylpropanoid metabolism and is suggested to respond in a feedback loop to changes in this defense-related class of compounds (Gaudinier et al., 2015).

METABOLITES AND STRESS RESPONSES

Metabolic responses to stress are ubiquitous and well described for many plant systems. Time-series experiments revealed that metabolic activities can respond to stress more quickly than transcriptional activities (Kaplan et al., 2007; Guy et al., 2008; Caldana et al., 2011; Fraire-Velázquez and Balderas-Hernández, 2013), thus making metabolic changes an important part of early stress responses. Several metabolites can directly influence plant stress responses (Rojas et al., 2014). A few important examples are discussed below.

Carbohydrate Metabolism

During freezing and drought, soluble sugars, such as sucrose, trehalose, fructans (fructose-based oligo- and polysaccharides), and the raffinose family of oligosaccharides can stabilize phospholipid membrane vesicles (Hincha et al., 2006; Livingston et al., 2009; Tarkowski and Van den Ende, 2015). Several sugars

emerged as important factors also during biotic stress signaling (Baena-González and Sheen, 2008; Figueroa and Lunn, 2016; Li and Sheen, 2016). Sucrose has been shown to regulate various stress-related responses including circadian clock genes, phytohormones, energy metabolism, cell wall and anthocyanin synthesis (Thibaud et al., 2004; Gomez-Ariza et al., 2007; Tognetti et al., 2013; Tauzin and Giardina, 2014). Glucose induces the pathogen defense proteins PR-1 and PR-5 in *Arabidopsis* via hexokinase1 (HXK1) signaling (Xiao et al., 2000; Moore et al., 2003; Cho et al., 2009). For fructose, a specific pathway has been proposed that involves abscisic acid and ethylene signaling (Cho and Yoo, 2011; Li et al., 2011). Engelsdorf et al. (2013) showed that carbohydrate availability influences defense against a hemibiotrophic fungus during its necrotrophic phase. The more carbohydrate is available, the better the plant defends. Similarly, increased relative fructose content enhances defense against the pathogen *Botrytis cinerea* in tomato (Lecompte et al., 2017). Besides the importance of sugars for stress signaling, plants organize sugar distribution in a way that pathogens have reduced access to carbohydrates. Local depletion of nutrients appears to cause a starvation effect that reduces pathogen propagation (Bezruczyk et al., 2018).

Amino Acid Pathways

In a 2010 paper, Liu and colleagues knocked out the amino acid transporter lht1 and correlated cellular depletion of the amino acid glutamine with altered redox status and more effective defense against several pathogens (Liu et al., 2010). These authors proposed a yet unknown negative effect of glutamine on defense signaling and a reduction in the pathogen's access to essential nutrients, similar to recent findings on *Pseudomonas*-primed systemic responses of *Arabidopsis* (Schwachtje et al., 2018). Stuttmann et al. (2011) described threonine as a potential growth inhibitor of the biotrophic oomycete *Hyaloperonospora arabidopsidis*, even though the underlying mechanism is yet unknown. It is suggested that indole-3-carboxylic acid activates defenses against *Plectosphaerella cucumerina* in *Arabidopsis* by inducing papillae deposition and H₂O₂ production, independently of salicylic acid and jasmonic acid (Gamir et al., 2014b). γ -Aminobutyric acid (GABA) interacts with quorum sensing of *Agrobacterium tumefaciens*, thus reducing pathogen virulence in tobacco (Chevrot et al., 2006). GABA also functions as a direct anti-herbivore defense in *Arabidopsis* (Scholz et al., 2015). Furthermore, the proline and the pyrroline-5-carboxylate (P5C) cycle are crucial for defense responses against pathogens and abiotic stresses (Liang et al., 2013; Qamar et al., 2015). Proline is involved in redox balance, osmoprotection, and stress signaling (Szabados and Savaure, 2009).

Polyamine Metabolism

Polyamines (e.g., spermine, spermidine, and putrescine) are aliphatic compounds that are synthesized from amino acids (e.g., arginine and ornithine). Polyamines are involved in many crucial processes of cell metabolism and the translation/transcription machinery, and induce ROS, Ca, and NO signaling (Alcázar et al., 2010). Salt, heat, and drought stress induce genes for

polyamine synthesis (Tiburcio et al., 2014; Liu et al., 2015; Miller-Fleming et al., 2015) and enhanced tolerance of abiotic stresses is correlated with elevated levels of polyamines (Alcázar et al., 2010). Putrescine induces abscisic acid synthesis at the transcriptional level during cold stress (Cuevas et al., 2008). Spermine appears to protect *Arabidopsis* from heat stress by increasing the expression of genes encoding heat shock proteins (Sagor et al., 2013). The pretreatment of tomato fruits with spermine before heat shock promoted an increase in expression of signal transduction genes (e.g., calmodulin, serine/threonine protein kinase) along with genes related to phytohormone pathways. Moreover, polyamines can modulate chromatin structure (Pasini et al., 2014). It is also suggested that the connected putrescine, GABA, and proline pathways play an important role during abiotic stresses (Shelp et al., 2012; Legocka et al., 2017).

External Application of Natural Compounds

A few metabolites have been shown to prime pathogen-induced stress responses when externally applied to plants, e.g., thiamine (Ahn et al., 2007), riboflavin (Zhang et al., 2009), quercetin (Jia et al., 2010), and hexanoic acid (Aranega-Bou et al., 2014). Supposedly, all have in common an activation of the redox system that supports stress signaling. In a recent study, fumarate and citrate applications were shown to induce priming against *Pseudomonas syringae* in *Arabidopsis*, in the case of fumarate without changes in classical defense-related genes and hormones (Balmer et al., 2018). Thereby, Balmer et al. (2018) confirmed an earlier observation of systemic fumarate priming upon first exposure to the bacterial pathogen (Schwachtje et al., 2018). Further, a broad induction of plant defense systems in *Arabidopsis* was demonstrated after the application of melatonin (Weeda et al., 2014).

ARE STRESS-RELATED METABOLITES STORED IN THE VACUOLE?

To exert a long-term effect that primes future stress responses, relevant metabolites must be stored in a way that prevents them from negatively interfering with metabolism during the recovery phase and that inhibits degradation. Accumulation of metabolites in chloroplasts, mitochondria, or in the cytosol would likely disturb core metabolic processes that are necessary for recovery (Plaxton, 2005). Possible circumventions would be a reversible conjugation that alters the chemical property of the metabolite, or storage in a membrane-enclosed cellular compartment, such as the vacuole. The vacuole can occupy more than 80% of the cell's volume and is involved in multiple critical cellular functions including storage of metabolites and modification of cytosolic metabolism according to physiological requirements (Martinoia et al., 2012). The tonoplast enclosing the vacuole contains many identified membrane proteins that are responsible for loading and unloading a diverse set of metabolites (Martinoia et al., 2012). These transporters are integrated in a larger cellular network that responds to physiological requirements and stress responses (Martinoia

et al., 2007; Pommerrenig et al., 2018). The carboxylic acids fumarate, malate, and citrate represent major components of the vacuolar metabolome. For example, malate is transported by two proteins, a specific anion channel (Hafke et al., 2003) and a solute carrier (Emmerlich et al., 2003). Many other primary metabolites, such as amino acids and sugars, are stored in the vacuole (Neuhaus, 2007; Tohge et al., 2011; Szekowka et al., 2013; Hedrich et al., 2015). Vacuoles are also involved in plant defenses against herbivores and pathogens by storing and sequestering toxic metabolites. Thus, specific analyses of vacuolar metabolite compositions, e.g., by non-aqueous fractionation (Klie et al., 2011), are required to unravel a possible long-term storage of stress-induced metabolites that represent metabolic imprints and may function as a metabolic memory.

EXAMPLES OF STRESS-INDUCED METABOLIC IMPRINTS

The functional analysis of stress-induced metabolic responses has been a long-standing focus of plant physiology. In contrast, metabolic imprints following stress have rarely received attention but can now be described and analyzed in detail by large-scale experiments that combine metabolic phenotyping with global screening of other system levels and physiological analyses (e.g., Hemme et al., 2014). The fragmented knowledge on recovery processes may result from the simplifying assumption that plants will revert to the identical initial endogenous state after a transient environmental perturbation. The resetting of the initial state is thought to alleviate the need to expend energy for the maintenance of the stress-adapted state. While this assumption may be correct for the majority of metabolites, the past perturbations may leave an imprint on metabolism that lasts longer than may be expected from a system level that is notorious for its extremely rapid fluctuations (Urbanczyk-Wochniak et al., 2005; Kim et al., 2011). In the following, we will review evidence of metabolic imprinting and functions of imprints for priming of systems in the context of abiotic and biotic stresses.

Abiotic Stress

Abiotic stresses are known to prime plant systems for an enhanced stress response to a recurrent stress. Recent reviews highlight abiotic stress priming of temperature, drought, and other factors (Bruce et al., 2007; Hinch and Zuther, 2014; Hilker et al., 2015). In the following, we will first highlight proline imprints that were observed in the context of various stresses before we address more stress-specific metabolic imprints.

Proline Imprints Are Caused by Various Abiotic Stresses

Proline accumulation is one of the most studied metabolic stress responses. Upon environmental stress, proline is mainly generated from glutamic acid in chloroplasts and increases up to 100-fold in plants (Liang et al., 2013). Proline has several functions during stress responses, e.g., as an osmoprotectant, antioxidant, molecular

chaperone to protect protein integrity, pH buffer, or in some cases it may serve as a carbon and nitrogen source during stress recovery. Proline also enhances enzyme activities, triggers gene expression, and modulates mitochondrial functions (Szabados and Savouré, 2009). By increasing ROS production in mitochondria *via* the electron transport chain, proline regulates processes that support cell survival or induce apoptosis (Liang et al., 2013). Suppression of proline catabolism, for example *via* reduction of proline dehydrogenase gene expression, enhances tolerance toward salt and drought stress (Ibragimova et al., 2012). In *Arabidopsis*, proline accumulates strongly during a 4-day drought phase and declines to initial levels during a subsequent 4-day recovery phase (Sharma and Verslues, 2010). In contrast, the drought-resistant *Periploca sepium* increases proline levels continuously during a similar 4-day drought stress but maintains a proline imprint during a 4-day recovery phase (An et al., 2013). Even after a 8-day recovery, the newly developed buds of *Periploca sepium* still contained twice as much proline as control plants. Proline is apparently also important for recovery of tobacco plants from drought stress by suppressing a senescence-related promoter (Vancova et al., 2012). In addition, the proline concentration in salt stress-resistant salt cress (*Thellungiella halophila*, renamed to *Thellungiella salsuginea*, and *Eutrema salsugineum*) is significantly higher than in *Arabidopsis* already under control conditions (Taji et al., 2004; Benina et al., 2013; Lee et al., 2016). In this case, high proline levels may serve as a constitutive stress adaptation of an extremophile plant. Also, in *T. halophila*, proline levels increased during a 3-day recovery from cold stress but not during the 3-day stress phase itself. The imprint of the proline pool is accompanied by other metabolites, such as 5-hydroxyproline and sucrose (Benina et al., 2013). In *Arabidopsis*, 3 days after de-acclimation from cold acclimation, proline levels were still elevated in leaves. More freeze-tolerant *Arabidopsis* accessions showed higher levels than susceptible accessions after 3 days of recovery (Zuther et al., 2015).

Drought

Several studies describe imprints of metabolite pools other than proline after exposure to drought. Primary metabolites, e.g., sugars and organic acids, as well as several secondary metabolites maintain a characteristic imprint in the resurrection plant *Haberlea rhodopensis* after 2 days of recovery from an 8-day drought period (Moyankova et al., 2014). A similar duration of drought stress and recovery, 8–10-day stress and 2-day recovery, causes a different metabolic imprint of *Medicago sativa* nodules (Naya et al., 2007). In this symbiotic system, pools of several primary metabolites remain reduced during drought recovery. A recent study describes the metabolic recovery of drought-stressed sugar beets (Wedeking et al., 2018). The authors found a transient normalization of most of the measured metabolites after 8 days of recovery from a 13-day drought stress period. Interestingly, during the following 4 days, several amino acids (e.g., phenylalanine, tyrosine, and leucine) again accumulated in leaves, indicating a metabolic stress imprint that may be beneficial for a subsequent second drought phase.

Low Temperature

Faster than drought stress, temperature stress may rapidly revert. Temperature changes cause metabolic responses that are particularly well characterized after heat shock or extended cold exposure (e.g., Kaplan et al., 2004, 2007; Guy et al., 2008). The temperature-induced metabolic responses comprise strong changes in a wide range of metabolite pools that indicate global reprogramming of primary metabolism in *Arabidopsis* rosettes. Data on metabolomic and transcriptomic cold recovery describe a 24-h metabolic imprint after 4 days of exposure to 4°C (Kaplan et al., 2004, 2007) and reveal several interesting aspects. Firstly, most of the cold-induced transcript changes returned to the pre-stress state after 24 h. In contrast, primary metabolism was only partially recovered. These differences between transcriptional and metabolic recovery from cold stress were recently confirmed in greater detail by Pagter et al. (2017). Secondly, the metabolite profile of the recovery phase differed significantly from all measured time points during cold exposure, supporting the observation that the metabolic reorganization after stress exhibits different kinetics than the stress response. Cold de-acclimating metabolism was associated with partially maintained enhanced freeze tolerance, which apparently remains active at least 3 days into cold-recovery (Kaplan et al., 2004; Zuther et al., 2015). Zuther et al. (2018) reported genetic differences in the transcriptomic and metabolic patterns during cold memory of the *Arabidopsis* ecotypes, Col-0 and N14.

A complex picture of metabolic reorganization during recovery was described after freezing stress of crown tissue of oat (*Avena sativa* L.) by Henson et al. (2014). After 3 weeks of cold acclimation and 1 day of freezing, plants were monitored during 14 days of recovery. At the end of recovery, several amino acids were largely increased compared to non-stressed plants, and several sugars and organic acids were reduced. Moreover, the metabolic profile differed markedly from what is observed after cold stress recovery, indicating that this overwintering species relies on specific regulations for freezing resistance.

Analyses of *Hordeum vulgare* also show stress-imprinted metabolites that are linked to frost tolerance (Mazucotelli et al., 2006). For example, 8-day-old barley seedlings were freeze-stressed at −3°C for 16 h and allowed to recover for 48 h at 22°C. This treatment resulted in 16-fold higher GABA levels at the end of the recovery phase. GABA and its precursor glutamate are part of the GABA-shunt that is linked to the tricarboxylic acid cycle where it bypasses two reaction steps from α -ketoglutarate to succinate. Besides glutamate, putrescine and proline can also be catabolized via GABA (Shelp et al., 2012; Signorelli et al., 2015). The GABA-shunt has a central role in carbon/nitrogen metabolism and stress signaling, for example for cell death promotion in response to pathogens or for cold tolerance (Mazucotelli et al., 2006; Fait et al., 2007; Michaeli and Fromm, 2015). However, the role of GABA during cold-/frost-stress and possible GABA pool imprints are still not fully understood.

High Temperature

Elevated temperatures leave metabolic imprints in photosynthetic microorganisms. A recent large-scale study describes the temporal

succession of heat stress responses of *Chlamydomonas reinhardtii* during a 24-h induction phase after shift from 25 to 42°C and the fate of system imprints during 8-h recovery at 25°C (Hemme et al., 2014). In this experiment, cell division stopped during heat treatment and remained so during the 8 h of recovery, resulting in measurements that represent the metabolic state of cells that all individually experienced the heat stress. Similar to the example of 4°C cold stress in *Arabidopsis* (Kaplan et al., 2004, Pagter et al., 2017), the metabolome, as well as the proteome, recovered only in part and retained imprints, regarding, e.g., TCA intermediates and sugar phosphates. Importantly, the pattern of metabolic induction again differed from the pattern of metabolic recovery.

Besides temperature perturbations, other abiotic stresses have been shown to generate lasting imprints. A 6-h oxidative stress that was induced by menadione generated an imprint on primary metabolism in *Arabidopsis* roots that lasted at least 30 h into recovery (Lehmann et al., 2012). GABA was part of this imprint, like proline and other amino acids which remained at a high level, as well as several sugars and sugar phosphates.

Biotic Stress

Biotic stresses are perhaps the best understood stresses regarding primed plant systems. Recent reviews highlight the importance of biotic stress priming for enhanced responses toward a broad range of insects and pathogens that negatively influence plant performance and crop production (Frost et al., 2008; Conrath et al., 2015; Hilker et al., 2015; Mauch-Mani et al., 2017). At the metabolite level, biotic priming is mainly studied with respect to volatile organic compounds, oviposition, and beneficial or pathogenic microorganisms that are associated with a plant and may prime systemic tissue. Even though it has often been shown that during biotic stresses, metabolism in local and systemic plant parts is severely affected (Schwachtje and Baldwin, 2006; Lemoine et al., 2013; Zhou et al., 2015), studies on persistent metabolic changes during and after recovery from pathogen or insect stress are rare. This applies specifically for interactions of plants with microorganisms, since these are continuously associated with the plant, either as beneficial root colonizers or as leaf pathogens, thus making a clearly defined recovery phase after a time-limited stress or induction phase unfeasible. Nevertheless, several metabolites have so far been associated with priming against biotic stresses.

Plant amino acid metabolism is well known to contribute to the priming of defense responses (Gamir et al., 2014a). For example, the lysine catabolite pipecolic acid can act as a key regulator of SAR (Návarová et al., 2012; Zeier, 2013). Several amino acids and intermediates of the TCA cycle are regulated during priming with pathogenic *Pseudomonas syringae* or the chemical β -aminobutyric acid, i.e., BABA (Pastor et al., 2014). The content of most amino acids was reduced in these experiments, but cysteine, methionine, tryptophan, and tyrosine were specifically induced by bacteria or BABA during 48 h. Fumarate and malate were induced by BABA. These two organic acids were also induced

in another study that investigated systemic priming-related effects of infection of *Arabidopsis* with *Pseudomonas syringae* (Schwachtje et al., 2018). A systemic increase of fumarate and malate was observed for 4 days, whereas the transcriptional profile did not explain the altered metabolite levels. This study suggested a lasting metabolic priming effect in systemic tissue that includes storage of metabolites, e.g., fumarate and malate. These metabolites may be readily available to support energy and carbon demands during a subsequent pathogen (*Pseudomonas syringae*) infection.

Altered amino acid levels after contact with a pathogen can have multiple and possibly conflicting functions. Several amino acids are precursors of important defense metabolites, e.g., alkaloids, phenylpropanoids, and glucosinolates. On the other hand, invasive pathogens like *Pseudomonas syringae* propagate in the apoplast and are exclusively dependent on extracellular plant metabolites. Reduction of sugars and amino acids in the apoplast and, as indicated by transcript changes, likely also other nitrogen resources should be an effective defense strategy that attenuates pathogen propagation and thereby increases the efficiency of other defense mechanisms (Seifi et al., 2013; Bezruczyk et al., 2018; Schwachtje et al., 2018). Several imprinted metabolic signals have been identified that may contribute to the modulation of SAR, including a glycerol-3-phosphate-dependent yet non-identified signal, azelaic acid, dehydroabietinal, jasmonic acid, and methyl salicylate (Dempsey and Klessig, 2012). The metabolic signals that are linked to SAR or other primed responses will yield intriguing novel insights into imprinted primary metabolism/energy status and the function of such imprints for efficiently primed plant responses.

In their natural environment, plants usually face more than one type of stress. The physiological responses toward various stress combinations, simultaneous or successive, have been addressed by recent studies (reviewed in Suzuki et al., 2014). The effects on plant performance can be synergistic, neutral, or conflicting (Crisp et al., 2016; Lawas et al., 2018) and it will be a demanding task to unravel how metabolite-based priming and priming in general by a certain stress may influence plant responses to other types of stress.

EXPERIMENTAL APPROACHES

The successful search for priming-related metabolites relies on the timing of experiments. Mostly, stress studies focus on the immediate response of the plant system toward an applied stress, but rarely focus on the long-term effects on plant metabolism. The recovery phase after a stress event is crucial for the establishment of priming and should thus be studied more extensively. The history of plants prior to stress experiments is rarely controlled and comparable between experiments. These methodology details must be described in detail to enhance reproducibility of stress experiments.

Several factors interfere with the metabolic state of a plant during the recovery phase and must be experimentally addressed.

As described above, the metabolic composition of plant tissues is an integral of perceived environmental stresses (Gratani, 2014) and may lead to variation among individual plants even under standardized conditions (e.g., Sanchez et al., 2010). Metabolism is continuously regulated by the circadian clock (Farre and Weise, 2012) and this regulation also affects stress responses themselves on genetic and metabolic levels (Lu et al., 2017). For example, glucosinolate accumulation follows the circadian rhythm in *Arabidopsis* and jasmonic acid-based defenses are synchronized with the likeliness of herbivore attack (Goodspeed et al., 2012). In return, several biotic stressors have recently been shown to influence the cycle length of the circadian clock, e.g., pathogens and insects (Sharma and Bhatt, 2015; Joo et al., 2018; Li et al., 2018). This requires experimental setups with extended sampling time points during the day. Also, the influence of the ontogenetic stage, i.e., the effects of endogenous physiological aging mechanisms, on induced metabolic responses should be addressed. Furthermore, as described above, priming-related metabolites may be stored in certain cell compartments (e.g., the vacuole). Subcellular localization of metabolites is difficult to assess but can be addressed by non-aqueous fractionation (Klie et al., 2011). Because temporal effects are essential for the assessment of induced, imprinted, and primed responses, care should be taken to design time-series experiments with extended and high temporal resolution including the coverage of diurnal changes. High replication is advised due to varying histories and developmental variation of individual plants (e.g., Peters et al., 2018), this applies particularly to field experiments. Further, the high chemodiversity of plant metabolites entails many different chemical properties. To find new candidates for priming, the application of multiple chromatography systems for untargeted metabolic profiling should be taken into account (e.g., Nakabayashi and Saito, 2015; Vasilev et al., 2016).

CONCLUSION

Recent publications tackle the study of metabolic imprints by analyses of recurrent perturbations or recovering plant systems and discover functions of novel primed metabolites and metabolic pathways (Gamir et al., 2014a; Balmer et al., 2015). Intensified research on the potential functions of metabolic imprints should be highly fruitful and yield novel insights into priming phenomena. This view is supported by recent findings that demonstrate surprisingly diverse effects of metabolites on stress metabolism, signaling, and transcription. The vast chemical diversity of plants will likely yield new candidates of metabolic regulation or priming.

Currently, the knowledge of the short- to long-term kinetics of metabolic imprints is fragmented. This fact renders vague the link between observed metabolic imprints and their potential function as priming signals or memory of past stress events. Except for the known signaling metabolites that are involved in primed plant responses, the nature,

characteristics, and role of metabolic imprinting or priming remain mostly unknown not least because stress recovery is rarely investigated in depth by studies employing modern large-scale metabolomic, proteomic, transcriptomic, or epigenetic tools. From advanced analyses of metabolic imprints, we expect to discover new priming mechanisms and to gain insight into the major contributions of metabolism to priming and potentially short-lived or even longer-lasting non-neural, cellular memory.

AUTHOR CONTRIBUTIONS

JS and JK developed the concepts and wrote the manuscript with contributions of all other co-authors.

REFERENCES

- Ahn, I. P., Kim, S., Lee, Y. H., and Suh, S. C. (2007). Vitamin B1-induced priming is dependent on hydrogen peroxide and the NPR1 gene in *Arabidopsis*. *Plant Physiol.* 143, 838–848. doi: 10.1104/pp.106.092627
- Alcázar, R., Altabella, T., Marco, F., Bortolotti, C., Reymond, M., Koncz, C., et al. (2010). Polyamines: molecules with regulatory functions in plant abiotic stress tolerance. *Planta* 231, 1237–1249. doi: 10.1007/s00425-010-1130-0
- An, Y., Zhang, M., Liu, G., Han, R., and Liang, Z. (2013). Proline accumulation in leaves of *Periploca sepium* via both biosynthesis up-regulation and transport during recovery from severe drought. *PLoS One* 8:e69942. doi: 10.1371/journal.pone.0069942
- Aranega-Bou, P., de la O Leyva, M., Finiti, I., Garcia-Agustin, P., and González-Bosch, C. (2014). Priming of plant resistance by natural compounds. Hexanoic acid as a model. *Front. Plant Sci.* 5:488. doi: 10.3389/fpls.2014.00488
- Baena-González, E., and Sheen, J. (2008). Convergent energy and stress signaling. *Trends Plant Sci.* 13, 474–482. doi: 10.1016/j.tplants.2008.06.006
- Baena-González, E. (2010). Energy signaling in the regulation of gene expression during stress. *Mol. Plant* 3, 300–313. doi: 10.1093/mp/ssp113
- Bäurle, I. (2016). Plant heat adaptation: priming in response to heat stress. *F1000Res.* 5:694. doi: 10.12688/f1000research.7526.1
- Balmer, A., Pastor, V., Gamir, J., Flors, V., and Mauch-Mani, B. (2015). The 'prime-ome': towards a holistic approach to priming. *Trends Plant Sci.* 20, 443–452. doi: 10.1016/j.tplants.2015.04.002
- Balmer, A., Pastor, V., Glauser, G., and Mauch-Mani, B. (2018). Tricarboxylates induce defense priming against bacteria in *Arabidopsis thaliana*. *Front. Plant Sci.* 9:1221. doi: 10.3389/fpls.2018.01221
- Bandoly, M., Grichnik, R., Hilker, M., and Steppuhn, A. (2016). Priming of anti-herbivore defence in *Nicotiana attenuata* by insect oviposition: herbivore-specific effects. *Plant Cell Environ.* 39, 848–859. doi: 10.1111/pce.12677
- Benina, M., Obata, T., Mehterov, N., Ivanov, I., Petrov, V., Toneva, V., et al. (2013). Comparative metabolic profiling of *Haberlea rhodopensis*, *Thellungiella halophylla*, and *Arabidopsis thaliana* exposed to low temperature. *Front. Plant Sci.* 4, 1–11. doi: 10.3389/fpls.2013.00499
- Bezruczyk, M., Yang, J., Eom, J.-S., Prior, M., Soss, D., Hartwig, T., et al. (2018). Sugar flux and signaling in plant-microbe interactions. *Plant J.* 93, 675–685. doi: 10.1111/tpl.13775
- Bolton, M. D. (2009). Primary metabolism and plant defense—fuel for the fire. *Mol. Plant-Microbe Interact.* 22, 487–497. doi: 10.1094/MPMI-22-5-0487
- Bonawitz, N. D., Soltan, W. L., Blatchley, M. R., Powers, B. L., Hurlock, A. K., Seals, L. A., et al. (2012). REF4 and RFR1, subunits of the transcriptional coregulatory complex mediator, are required for phenylpropanoid homeostasis in *Arabidopsis*. *J. Biol. Chem.* 287, 5434–5445. doi: 10.1074/jbc.M111.312298
- Bruce, T. J. A., Matthes, M. C., Napier, J. A., and Pickett, J. A. (2007). Stressful “memories” of plants: evidence and possible mechanisms. *Plant Sci.* 173, 603–608. doi: 10.1016/j.plantsci.2007.09.002

FUNDING

We acknowledge the Max-Planck Society and the German Research Foundation (DFG) for funding the Collaborative Research Centre 973 “Priming and Memory of Organismic Responses to Stress” (www.sfb973.de). We thank the National Council for the Improvement of Higher Education—CAPES of Brazil—for the scholarship provided to AF.

ACKNOWLEDGMENTS

We acknowledge the long-standing support by Prof. Dr. L. Willmitzer, Prof. Dr. M. Stitt, and Prof. Dr. R. Bock (Max-Planck-Institute of Molecular Plant Physiology, Potsdam, Germany).

- Caldana, C., Degenkolbe, T., Cuadros-Inostroza, A., Klie, S., Sulpice, R., Leisse, A., et al. (2011). High-density kinetic analysis of the metabolomic and transcriptomic response of *Arabidopsis* to eight environmental conditions. *Plant J.* 67, 869–884. doi: 10.1111/j.1365-3113.2011.04640.x
- Chevrot, R., Rosen, R., Haudecoer, E., Cirou, A., Shelp, B. J., Ron, E., et al. (2006). GABA controls the level of quorum-sensing signal in *Agrobacterium tumefaciens*. *PNAS* 103, 7460–7464. doi: 10.1073/pnas.0600313103
- Cho, J. I., Ryoo, N., Eom, J. S., Lee, D. W., Kim, H. B., Jeong, S. W., et al. (2009). Role of the rice hexokinases OsHXX5 and OsHXX6 as glucose sensors. *Plant Physiol.* 149, 745–759. doi: 10.1104/pp.108.131227
- Cho, Y. H., and Yoo, S. D. (2011). Signaling role of fructose mediated by FINS1/FBP in *Arabidopsis thaliana*. *PLoS Genet.* 7:e1001263. doi: 10.1371/journal.pgen.1001263
- Chubukov, V., Uhr, M., Le Chat, L., Kleijn, R. J., Jules, M., Link, H., et al. (2013). Transcriptional regulation is insufficient to explain substrate-induced flux changes in *Bacillus subtilis*. *Mol. Syst. Biol.* 9:709. doi: 10.1038/msb.2013.66
- Conrath, U., Beckers, G. J. M., Langenbach, C. J. L., and Jaskiewicz, M. R. (2015). Priming for enhanced defense. *Annu. Rev. Phytopathol.* 53, 97–119. doi: 10.1146/annurev-phyto-080614-120132
- Crisp, P. A., Ganguly, D., Eichten, S. R., Borevitz, J. O., and Pogson, B. J. (2016). Reconsidering plant memory: Intersections between stress recovery, RNA turnover, and epigenetics. *Sci. Adv.* 2:e1501340. doi: 10.1126/sciadv.1501340
- Cuevas, J. C., Lopez-Cobollo, R., Alcázar, R., Zarza, X., Koncz, C., Altabella, T., et al. (2008). Putrescine is involved in *Arabidopsis* freezing tolerance and cold acclimation by regulating abscisic acid levels in response to low temperature. *Plant Physiol.* 148, 1094–1105. doi: 10.1104/pp.108.122945
- Dangl, J. L., and Jones, J. D. G. (2001). Plant pathogens and integrated defence responses to infection. *Nature* 411, 826–833. doi: 10.1038/35081161
- De Deyn, B. G., Cornelissen, J. H. C., and Bardgett, R. D. (2008). Plant functional traits and soil carbon sequestration in contrasting biomes. *Ecol. Lett.* 11, 516–531. doi: 10.1111/j.1461-0248.2008.01164.x
- Dempsey, D. M. A., and Klessig, D. F. (2012). SOS – too many signals for systemic acquired resistance? *Trends Plant Sci.* 17, 538–545. doi: 10.1016/j.tplants.2012.05.011
- Ding, Y., Fromm, M., and Avramova, Z. (2012). Multiple exposures to drought ‘train’ transcriptional responses in *Arabidopsis*. *Nat. Commun.* 3:740. doi: 10.1038/ncomms1732
- Dobrenel, T., Caldana, C., Hanson, J., Robaglia, C., Vincentz, M., Veit, B., et al. (2016). TOR signaling and nutrient sensing. *Annu. Rev. Plant Biol.* 67, 261–285. doi: 10.1146/annurev-arplant-043014-114648
- Donohue, K. (2009). Completing the cycle: maternal effects as the missing link in plant life histories. *Philos. Trans. R. Soc. B* 364, 1059–1074. doi: 10.1098/rstb.2008.0291
- Emmerlich, V., Linka, N., Reinhold, T., Hurth, M. A., Traub, M., Martinoia, E., et al. (2003). The plant homolog to the human sodium/dicarboxylic cotransporter

- is the vacuolar malate carrier. *PNAS* 100, 11122–11126. doi: 10.1073/pnas.1832002100
- Engelberth, J., Alborn, H. T., Schmelz, E. A., and Tumlinson, J. H. (2004). Airborne signals prime plants against insect herbivore attack. *Proc. Natl. Acad. Sci. U. S. A.* 101, 1781–1785. doi: 10.1073/pnas.0308037100
- Engelsdorf, T., Horst, R. J., Pröls, R., Pröschel, M., Dietz, F., Hükelhoven, R., et al. (2013). Reduced Carbohydrate Availability Enhances the Susceptibility of Arabidopsis toward *Colletotrichum higginsianum*. *Plant Physiol.* 162, 225–238. doi: 10.1104/pp.112.209676
- Fait, A., Fromm, H., Walter, D., Galili, G., and Fernie, A. F. (2007). Highway or byway: the metabolic role of the GABA shunt in plants. *Trends Plant Sci.* 13, 14–19. doi: 10.1016/j.tplants.2007.10.005
- Farre, E. V., and Weise, S. E. (2012). The interactions between the circadian clock and primary metabolism. *Curr. Opin. Plant Biol.* 15, 293–300. doi: 10.1016/j.pbi.2012.01.013
- Figueroa, C. M., and Lunn, J. E. (2016). A tale of two sugars: trehalose 6-phosphate and sucrose. *Plant Physiol.* 172, 7–27. doi: 10.1104/pp.16.00417
- Fraire-Velázquez, S., and Balderas-Hernández, V. E. (2013). “Abiotic stress in plants and metabolic responses” in *Abiotic Stress - Plant Responses and Applications in Agriculture*. eds. K. Vahdati and C. Leslie (Rijeka: InTech). doi: 10.5772/45842
- Francisco, M., Joseph, B., Caligagan, H., Li, B., Corwin, J. A., Lin, C., et al. (2016). The defense metabolite, allyl glucosinolate, modulates *Arabidopsis thaliana* biomass dependent upon the endogenous glucosinolate pathway. *Front. Plant Sci.* 7:774. doi: 10.3389/fpls.2016.00774
- Frost, C. J., Mescher, M. C., Carlson, J. E., and De Moraes, C. M. (2008). Plant defense priming against herbivores: getting ready for a different battle. *Plant Physiol.* 146, 818–824. doi: 10.1104/pp.107.113027
- Gamir, J., Sanchez-Bel, P., and Flors, V. (2014a). Molecular and physiological stages of priming: how plants prepare for environmental challenges. *Plant Cell Rep.* 33, 1935–1949. doi: 10.1007/s00299-014-1665-9
- Gamir, J., Pastor, V., Kaefer, A., Cerezo, M., and Flors, V. (2014b). Targeting novel chemical and constitutive primed metabolites against *Plectosphaerella cucumerina*. *Plant J.* 78, 227–240. doi: 10.1111/tpl.12465
- Gaudinier, A., Tang, M., and Kliebenstein, D. J. (2015). Transcriptional networks governing plant metabolism. *Curr. Plant Biol.* 3–4, 56–64. doi: 10.1016/j.cpb.2015.07.002
- Gomez-Ariza, J., Campo, S., Rufat, M., Estopa, M., Messeguer, J., San Segundo, B., et al. (2007). Sucrose-mediated priming of plant defense responses and broad-spectrum disease resistance by overexpression of the maize pathogenesis-related PRms protein in rice plants. *Mol. Plant-Microbe Interact.* 20, 832–842. doi: 10.1094/MPMI-20-7-0832
- González, A., and Hall, M. N. (2017). Nutrient sensing and TOR signaling in yeast and mammals. *EMBO J.* 36, 397–408. doi: 10.15252/embj.201696010
- Goodspeed, D., Chehab, E. W., Min-Venditti, A., Braam, J., and Covington, M. F. (2012). Arabidopsis synchronizes jasmonate-mediated defense with insect circadian behavior. *PNAS* 109, 4674–4677. doi: 10.1073/pnas.1116368109
- Gratani, L. (2014). Plant phenotypic plasticity in response to environmental factors. *Adv. Bot.* 2014:17. doi: 10.1155/2014/208747
- Guy, C. L., Kaplan, F., Kopka, J., Selbig, J., and Hinch, D. K. (2008). Metabolomics of temperature stress. *Physiol. Plant.* 132, 220–235. doi: 10.1111/j.1399-3054.2007.00999.x
- Guy, C. L. (2003). Freezing tolerance of plants: current understanding and selected emerging concepts. *Can. J. Bot.* 81, 1216–1223. doi: 10.1139/b03-130
- Hafke, J. B., Hafke, Y., Smith, J. A. C., Lüttge, U., and Thiel, G. (2003). Vacuolar malate uptake is mediated by an anion-selective inward rectifier. *Plant J.* 35, 116–128. doi: 10.1046/j.1365-3113X.2003.01781.x
- Hawkes, C. V., and Sullivan, J. J. (2001). The impact of herbivory on plants in different resource conditions: a meta-analysis. *Ecology* 82, 2045–2058. doi: 10.1890/0012-9658(2001)082[2045:TIOHOP]2.0.CO;2
- Hedrich, R., Sauer, N., and Neuhaus, H. E. (2015). Sugar transport across the plant vacuolar membrane: nature and regulation of carrier proteins. *Curr. Opin. Plant Biol.* 25, 63–70. doi: 10.1016/j.pbi.2015.04.008
- Hemme, D., Veyel, D., Mühlhaus, T., Sommer, F., Jüppner, J., Unger, A. -K., et al. (2014). Systems-wide analysis of acclimation responses to long-term heat stress and recovery in the photosynthetic model organism *Chlamydomonas reinhardtii*. *Plant Cell* 26, 4270–4297. doi: 10.1105/tpc.114.130997
- Henson, C. A., Duke, S. H., and Livingston, D. P. (2014). Metabolic changes in *Avena sativa* crowns recovering from freezing. *PLoS One* 9:e93085. doi: 10.1371/journal.pone.0093085
- Hey, S. J., Byrne, E., and Halford, N. G. (2010). The interface between metabolic and stress signaling. *Ann. Bot.* 105, 197–203. doi: 10.1093/aob/mcp285
- Hilker, M., Schwachtje, J., Baier, M., Balazadeh, S., Bäurle, I., Geiselhardt, S., et al. (2015). Priming and memory of stress responses in organisms lacking a nervous system. *Biol. Rev.* 91, 1118–1133. doi: 10.1111/brv.12215
- Hinch, D. K., and Zuther, E. (2014). “Introduction: plant cold acclimation and freezing tolerance” in *Methods in molecular biology*. eds. D. K. Hinch and E. Zuther, Vol. 1166 (New York: Springer), 1–6.
- Hinch, D. K., Popova, A. V., and Cacela, C. (2006). “Effects of sugars on the stability and structure of lipid membranes during drying” in *Advances in planar lipid bilayers and liposomes*. ed. A. Leitmannova Liu, Vol. 3 (Amsterdam, The Netherlands: Elsevier), 189–217. doi: 10.1016/S1554-4516(05)03006-1
- Ibragimova, S. S., Kolodyazhnaya, Y. S., Gerasimova, S. V., and Kochetov, A. V. (2012). Partial suppression of gene encoding proline dehydrogenase enhances plant tolerance to various abiotic stresses. *Russ. J. Plant Physiol.* 59, 88–96. doi: 10.1134/S1021443712010086
- Jakab, G., Cottier, V., Toquin, V., Rigoli, G., Zimmerli, L., Métraux, J. P., et al. (2001). β -Aminobutyric acid-induced resistance in plants. *Eur. J. Plant Pathol.* 107, 29–37. doi: 10.1023/A:1008730721037
- Jaskiewicz, M., Conrath, U., and Peterhansel, C. (2011). Chromatin modification acts as a memory for systemic acquired resistance in the plant stress response. *EMBO Rep.* 12, 50–55. doi: 10.1038/embor.2010.186
- Jia, Z. H., Zou, B. H., Wang, X. M., Qiu, J. A., Ma, H., Gou, Z. H., et al. (2010). Quercetin-induced H₂O₂ mediates the pathogen resistance against *Pseudomonas syringae* pv. Tomato DC3000 in *Arabidopsis thaliana*. *Biochem. Biophys. Res. Commun.* 396, 522–527. doi: 10.1016/j.bbrc.2010.04.131
- Jisha, K. C., Vijayakumari, K., and Puthur, J. T. (2013). Seed priming for abiotic stress tolerance: an overview. *Acta Physiol. Plant.* 35, 1381–1396. doi: 10.1007/s11738-012-1186-5
- Joo, Y., Schuman, M. C., Goldberg, J. K., Wissgott, A., Kim, S. G., and Baldwin, I. T. (2018). Herbivory elicits changes in green leaf volatile production via jasmonate signaling and the circadian clock. *Plant Cell Environ.* 41, 2577–2588. doi: 10.1111/pce.13474
- Kaplan, F., Kopka, J., Haskell, D. W., Zhao, W., Schiller, K. C., Gatzke, N., et al. (2004). Exploring the temperature-stress metabolome of Arabidopsis. *Plant Physiol.* 136, 4159–4168. doi: 10.1104/pp.104.052142
- Kaplan, F., Kopka, J., Sung, D. Y., Zhao, W., Popp, M., Porat, R., et al. (2007). Transcript and metabolite profiling during cold acclimation of Arabidopsis reveals an intricate relationship of cold-regulated gene expression with modifications in metabolite content. *Plant J.* 50, 967–981. doi: 10.1111/j.1365-3113X.2007.03100.x
- Katz, E., Nisani, S., Yadav, B. S., Woldemariam, M. G., Shai, B., Obolski, U., et al. (2015). The glucosinolate breakdown product indole-3-carbinol acts as an auxin antagonist in roots of *Arabidopsis thaliana*. *Plant J.* 82, 547–555. doi: 10.1111/tpl.12824
- Kim, S. G., Jon, F., Gaquerel, E., Gulati, J., and Baldwin, I. T. (2011). Tissue specific diurnal rhythms of metabolites and their regulation during herbivore attack in a native tobacco, *Nicotiana attenuata*. *PLoS One* 6:e26214. doi: 10.1371/journal.pone.0026214
- Klie, S., Krueger, S., Krall, L., Giavalisco, P., Flügge, U. I., Willmitzer, L., et al. (2011). Analysis of the compartmentalized metabolome—a validation of the non-aqueous fractionation technique. *Front. Plant Sci.* 2:55. doi: 10.3389/fpls.2011.00055
- Krasensky, J., and Jonak, C. (2012). Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks. *J. Exp. Bot.* 63, 1593–1608. doi: 10.1093/jxb/err460
- Lawas, L. M. F., Zuther, E., Jagdish, S. V. K., and Hinch, D. K. (2018). Molecular mechanisms of combined heat and drought stress resilience in cereals. *Curr. Opin. Plant Biol.* 45, 212–217. doi: 10.1016/j.pbi.2018.04.002
- Lecompte, F., Nicot, P. C., Ripoll, J., Abro, M. A., Raimbault, A. K., Lopez-Lauri, F., et al. (2017). Reduced susceptibility of tomato stem to the necrotrophic fungus *Botrytis cinerea* is associated with a specific adjustment of fructose content in the host sugar pool. *Ann. Bot.* 119, 931–943. doi: 10.1093/aob/mcw240

- Lee, Y. P., Funk, C., Erban, A., Kopka, J., Köhl, K. I., Zuther, E., et al. (2016). Salt stress responses in a geographically diverse collection of *Eutrema/Thellungiella* spp. Accessions. *Funct. Plant Biol.* 43, 590–606. doi: 10.1071/FP15285
- Legocka, J., Sobieszczuk-Nowicka, E., Ludwicki, D., and Lehmann, T. (2017). Putrescine catabolism via DAO contributes to proline and GABA accumulation in roots of lupine seedlings growing under salt stress. *Acta Soc. Bot. Pol.* 86:3549. doi: 10.5586/asbp.3549
- Lehmann, M., Laxa, M., Sweetlove, L. J., Fernie, A. R., and Obata, T. (2012). Metabolic recovery of *Arabidopsis thaliana* roots following cessation of oxidative stress. *Metabolomics* 8, 143–153. doi: 10.1007/s11306-011-0296-1
- Lemoine, R., La Camera, S., Atanassova, R., Dédaldéchamp, F., Allario, T., Pourtau, N., et al. (2013). Source-to-sink transport of sugar and regulation by environmental factors. *Front. Plant Sci.* 4:272. doi: 10.3389/fpls.2013.00272
- Li, P., Wind, J. J., Shi, X., Zhang, H., Hanson, J., Smeeckens, S. C., et al. (2011). Fructose sensitivity is suppressed in *Arabidopsis* by the transcription factor ANAC089 lacking the membrane-bound domain. *Proc. Natl. Acad. Sci. U. S. A.* 108, 3436–3441. doi: 10.1073/pnas.1018665108
- Li, L., and Sheen, J. (2016). Dynamic and diverse sugar signalling. *Curr. Opin. Plant Biol.* 33, 116–125. doi: 10.1016/j.pbi.2016.06.018
- Li, Z., Bonaldi, K., Uribe, F., and Pruneda-Paz, J. L. (2018). A localized *Pseudomonas syringae* infection triggers systemic clock responses in *Arabidopsis*. *Curr. Biol.* 28, 630–639. doi: 10.1016/j.cub.2018.01.001
- Liang, X., Zhang, L., Natarajan, S. K., and Becker, D. F. (2013). Proline mechanisms of stress survival. *Antioxid. Redox Signal.* 19, 998–1011. doi: 10.1089/ars.2012.5074
- Liu, G., Ji, Y., Bhuiyan, N. H., Pilot, G., Selvaraj, G., Zou, J., et al. (2010). Amino acid homeostasis modulates salicylic acid-associated redox status and defense responses in *Arabidopsis*. *Plant Cell* 22, 3845–3863. doi: 10.1105/tpc.110.079392
- Liu, J. -H., Wang, W., Wu, H., Gong, X., and Moriguchi, T. (2015). Polyamines function in stress tolerance: from synthesis to regulation. *Frontiers in Plant Science* 6:827. doi: 10.3389/fpls.2015.00827
- Livingston, D. P. 3rd., Hinch, D. K., and Heyer, A. G. (2009). Fructan and its relationship to abiotic stress tolerance in plants. *Cell. Mol. Life Sci.* 66, 2007–2023. doi: 10.1007/s00018-009-0002-x
- Lu, H., McClung, R., and Zhang, C. (2017). Tick tock: circadian regulation of plant innate immunity. *Annu. Rev. Phytopathol.* 55, 287–311. doi: 10.1146/annurev-phyto-080516-035451
- Mackie, A., Keseler, I. M., Nolan, L., Karp, P. D., and Paulsen, I. T. (2013). Dead end metabolites—defining the known unknowns of the *E. coli* metabolic network. *PLoS One* 8:e75210. doi: 10.1371/journal.pone.0075210
- Malinovski, F. G., Thomson, M.-L. F., Nintemann, S. J., Jagd, L. M., Bourguine, B., Burrow, M., et al. (2017). An evolutionarily young defense metabolite influences the root growth of plants via the ancient TOR signaling pathway. *elife* 6:e29353. doi: 10.7554/eLife.29353
- Mauch-Mani, B., Baccelli, I., Luna, E., and Flors, V. (2017). Defense priming: an adaptive part of induced resistance. *Annu. Rev. Plant Biol.* 68, 485–512. doi: 10.1146/annurev-arplant-042916-041132
- Martinoia, E., Maeshima, M., and Neuhaus, H. E. (2007). Vacuolar transporters and their essential role in plant metabolism. *J. Exp. Bot.* 58, 83–102. doi: 10.1093/jxb/erl183
- Martinoia, E., Meyer, S., De Angeli, A., and Nagy, R. (2012). Vacuolar transporters in their physiological context. *Annu. Rev. Plant Biol.* 63, 183–213. doi: 10.1146/annurev-arplant-042811-105608
- Mazucotelli, E., Tartari, A., Cattivelli, L., and Forlani, G. (2006). Metabolism of γ -aminobutyric acid during cold acclimation and freezing and its relationship to frost tolerance in barley and wheat. *J. Exp. Bot.* 57, 3755–3766. doi: 10.1093/jxb/erl141
- Michaeli, S., and Fromm, H. (2015). Closing the loop on the GABA shunt in plants: are GABA metabolism and signalling entwined? *Front. Plant Sci.* 4:419. doi: 10.3389/fpls.2015.00419
- Miller-Fleming, L., Olin-Sandoval, V., Campbell, K., and Ralser, M. (2015). Remaining Mysteries of Molecular Biology: The Role of Polyamines in the Cell. *J. Mol. Biol.* 427, 3389–3406. doi: 10.1016/j.jmb.2015.06.020
- Moore, B., Zhou, L., Rolland, F., Hall, Q., Cheng, W. H., Liu, Y. X., et al. (2003). Role of the *Arabidopsis* glucose sensor HXK1 in nutrient, light, and hormonal signaling. *Science* 300, 332–336. doi:10.1126/science.1080585
- Moyankova, D., Mladenov, P., Berkov, S., Peshev, D., Georgieva, D., and Djilianov, D. (2014). Metabolic profiling of the resurrection plant *Haberlea rhodopensis* during desiccation and recovery. *Physiol. Plant.* 152, 675–687. doi: 10.1111/pp.12212
- Nakabayashi, R., and Saito, K. (2015). Integrated metabolomics for abiotic stress responses in plants. *Curr. Opin. Plant Biol.* 24, 10–16. doi: 10.1016/j.pbi.2015.01.003
- Návarová, H., Bernsdorff, F., Döring, A. C., and Zeier, J. (2012). Pipecolic acid, an endogenous mediator of defense amplification and priming, is a critical regulator of inducible plant immunity. *Plant Cell* 24, 5123–5141. doi: 10.1105/tpc.112.103564
- Naya, L., Ladrera, R., Ramos, J., González, E. M., Arrese-Igor, C., Minchin, F. R., et al. (2007). The response of carbon metabolism and antioxidant defenses of alfalfa nodules to drought stress and to the subsequent recovery of plants. *Plant Physiol.* 144, 1104–1114. doi: 10.1104/pp.107.099648
- Neuhaus, H. E. (2007). Transport of primary metabolites across the plant vacuolar membrane. *FEBS Lett.* 581, 2223–2226. doi: 10.1016/j.febslet.2007.02.003
- Onwuka, B., and Mang, B. (2018). Effects of soil temperature on some soil properties and plant growth. *Adv. Plants Agric. Res.* 8, 34–37. doi: 10.15406/apar.2018.08.00288
- Pagter, M., Alpers, J., Erban, A., Kopka, J., Zuther, E., and Hinch, D. K. (2017). Rapid transcriptional and metabolic regulation of the deacclimation process in cold acclimated *Arabidopsis thaliana*. *BMC Genomics* 18:731. doi: 10.1186/s12864-017-4126-3
- Pasini, A., Caldarella, C. M., and Giordano, E. (2014). Chromatin remodeling by polyamines and polyamine analogs. *Amino Acids* 46, 595–603. doi: 10.1007/s00726-013-1550-9
- Pastor, V., Balmer, A., Gamir, J., Flors, F., and Mauch-Mani, B. (2014). Preparing to fight back: generation and storage of priming compounds. *Front. Plant Sci.* 5:295. doi: 10.3389/fpls.2014.00295
- Peters, K., Worrlich, A., Weinhold, A., Alka, O., Balcke, G., Birkemeyer, C., et al. (2018). Current challenges in plant eco-metabolomics. *Int. J. Mol. Sci.* 19, 1385–1422. doi: 10.3390/ijms19051385
- Pinson, B., Vaur, S., Sagot, I., Couplier, F., Lemoine, S., and Daignan-Fornier, B. (2009). Metabolic intermediates selectively stimulate transcription factor interaction and modulate phosphate and purine pathways. *Genes Dev.* 23, 1399–1407. doi: 10.1101/gad.521809
- Plaxton, W. C. (2005). “Principles of metabolic control” in *Functional metabolism: regulation and adaptation* 1. ed. K. B. Storey, Vol. 1 (New Jersey: Wiley). doi: 10.1002/047167558X.ch1
- Pommerrenig, B., Ludewig, F., Cvetkovic, J., Trentmann, O., Klemens, P. A. W., and Neuhaus, H. E. (2018). In concert: orchestrated changes in carbohydrate homeostasis are critical for plant abiotic stress tolerance. *Plant Cell Physiol.* 59, 1290–1299. doi: 10.1093/pcp/pcy037
- Pozo, M. J., Verhage, A., García-Andrade, J., García, J. M., and Azcón-Aguilar, C. (2009). “Priming plant defences against pathogens by arbuscular mycorrhizal fungi” in *Mycorrhizas: functional processes and ecological impact*. eds. C. Azcón-Aguilar, J. M. Barea, S. Gianinazzi, and V. Gianinazzi-Pearson (Berlin and Heidelberg: Springer), 137–149. doi: 10.1007/978-3-540-87978-7_9
- Qamar, A., Mysore, K. S., and Senthil-Kumar, M. (2015). Role of proline and pyrroline-5-carboxylate metabolism in plant defense against invading pathogens. *Front. Plant Sci.* 6:503. doi: 10.3389/fpls.2015.00503
- Rasmann, S., De Vos, M., Casteel, C. L., Tian, D., Halitschke, R., Sun, J. Y., et al. (2011). Herbivory in the previous generation primes plants for enhanced insect resistance. *Plant Physiol.* 158, 854–863. doi: 10.1104/pp.111.187831
- Rexin, D., Meyer, C., Robaglia, C., and Veit, B. (2015). TOR signalling in plants. *Biochem. J.* 470, 1–14. doi: 10.1042/BJ20150505
- Rojas, C. M., Senthil-Kumar, M., Tzin, V., and Mysore, K. S. (2014). Regulation of primary plant metabolism during plant-pathogen interactions and its contributions to plant defense. *Front. Plant Sci.* 5:17. doi: 10.3389/fpls.2014.00017
- Sagor, G. H. M., Berberich, T., Takahashi, Y., Niitsu, M., and Kusano, T. (2013). The polyamine spermine protects *Arabidopsis* from heat stress-induced damage by increasing expression of heat shock-related genes. *Transgenic Res.* 22, 595–605. doi: 10.1007/s11248-012-9666-3
- Sampaio, L., Edrada-Ebel, R., and Da Costa, F. B. (2015). Effect of the environment on the secondary metabolic profile of *Tithonia diversifolia*: a model for

- environmental metabolomics of plants. *Sci. Rep.* 6:29265. doi: 10.1038/srep29265
- Sanchez, D. H., Szymanski, J., Erban, A., Udvardi, M. K., and Kopka, J. (2010). Mining for robust transcriptional and metabolic responses to long-term salt stress: a case study on the model legume *Lotus japonicus*. *Plant Cell Environ.* 33, 468–480. doi: 10.1111/j.1365-3040.2009.02047.x
- Sani, E., Herzyk, P., Perrella, G., Colot, V., and Amtmann, A. (2013). Hyperosmotic priming of *Arabidopsis* seedlings establishes a long-term somatic memory accompanied by specific changes of the epigenome. *Genome Biol.* 14:R59. doi: 10.1186/gb-2013-14-6-r59
- Scholz, S. S., Reichelt, M., Mekonnen, D. W., Ludewig, F., and Mithöfer, A. (2015). Insect herbivory-elicited gaba accumulation in plants is a wound-induced, direct, systemic, and jasmonate-independent defense response. *Front. Plant Sci.* 6:1128. doi: 10.3389/fpls.2015.01128
- Schuman, M. C., and Baldwin, I. T. (2015). The layers of plant responses to herbivores. *Annu. Rev. Entomol.* 61, 373–394. doi: 10.1146/annurev-ento-010715-023851
- Schwachtje, J., and Baldwin, I. T. (2006). Why does herbivore attack reconfigure primary metabolism? *Plant Physiol.* 146, 845–851. doi: 10.1104/pp.107.112490
- Schwachtje, J., Fischer, A., Erban, A., and Kopka, J. (2018). Primed primary metabolism in systemic leaves: a functional systems analysis. *Sci. Rep.* 8:216. doi: 10.1038/s41598-017-18397-5
- Schwender, J., König, C., Klapperstück, M., Heinzel, N., Munz, E., Hebbelmann, I., et al. (2014). Transcript abundance on its own cannot be used to infer fluxes in central metabolism. *Front. Plant Sci.* 5:668. doi: 10.3389/fpls.2014.00668
- Seip, B., and Innis, C. A. (2016). How widespread is metabolite sensing by ribosome-arresting nascent peptides? *J. Mol. Biol.* 428, 2217–2227. doi: 10.1016/j.jmb.2016.04.019
- Seifi, H. S., Van Bockhaven, J., Angenon, G., and Höfte, M. (2013). Glutamate metabolism in plant disease and defense: friend or foe? *Mol. Plant-Microbe Interact.* 26, 475–485. doi: 10.1094/MPMI-07-12-0176-CR
- Sharma, S., and Verslues, P. E. (2010). Mechanisms independent of abscisic acid (ABA) or proline feedback have a predominant role in transcriptional regulation of proline metabolism during low water potential and stress recovery. *Plant Cell Environ.* 33, 1838–1851. doi: 10.1111/j.1365-3040.2010.02188.x
- Sharma, M., and Bhatt, D. (2015). The circadian clock and defence signalling in plants. *Mol. Plant Pathol.* 16, 210–218. doi: 10.1111/mpp.12178
- Shellhammer, J. P., Morin-Kensicki, E., Matson, J. P., Yin, G., Isom, D. G., Campbell, S., et al. (2017). Amino acid metabolites that regulate G protein signaling during osmotic stress. *PLoS Genet.* 13:e1006829. doi: 10.1371/journal.pgen.1006829
- Shelp, B. J., Bozzo, G. G., Trobacher, C. P., Zarei, A., Deyman, K. L., and Brikis, C. J. (2012). Hypothesis/review: Contribution of putrescine to 4-aminobutyrate (GABA) production in response to abiotic stress. *Plant Sci.* 193–194, 130–135. doi: 10.1016/j.plantsci.2012.06.001
- Shen, Y., Issakidis-Bourguet, E., and Zhou, D.-X. (2016). Perspectives on the interactions between metabolism, redox, and epigenetics in plants. *J. Exp. Bot.* 67, 5291–5300. doi: 10.1093/jxb/erw310
- Signorelli, S., Dans, P. D., Coitino, E. L., Borsani, O., and Monza, J. (2015). Connecting proline and α -aminobutyric acid in stressed plants through non-enzymatic reactions. *PLoS One* 10:e0115349. doi: 10.1371/journal.pone.0115349
- Stief, A., Altmann, S., Hoffmann, K., Pant, B. D., Scheible, W. R., and Bäurle, I. (2014). *Arabidopsis* *miR156* regulates tolerance to recurring environmental stress through SPL transcription factors. *Plant Cell* 26, 1792–1807. doi: 10.1105/tpc.114.123851
- Stuttman, J., Hubberten, H.-M., Rietz, S., Kaur, J., Muskett, P., and Guerois, R., et al. (2011). Perturbation of *Arabidopsis* amino acid metabolism causes incompatibility with the adapted biotrophic pathogen *Hyaloperonospora arabidopsidis*. *Plant Cell* 23, 2788–2803. doi: 10.1105/tpc.111.087684
- Sultan, S. E. (2000). Phenotypic plasticity for plant development, function and life history. *Trends Plant Sci.* 5, 537–542. doi: 10.1016/S1360-1385(00)01797-0
- Suzuki, N., Rivero, R. M., Shulaev, V., Blumwald, E., and Mittler, R. (2014). Abiotic and biotic stress combinations. *New Phytol.* 203, 32–43. doi: 10.1111/nph.12797
- Suzuki, N., Koussevitzky, S., Mittler, R., and Miller, G. (2011). ROS and redox signalling in the responses of plants to abiotic stress. *Plant Cell Environ.* 35, 259–270. doi: 10.1111/j.1365-3040.2011.02336.x
- Szabados, L., and Savouré, A. (2009). Proline: a multifunctional amino acid. *Trends Plant Sci.* 15, 89–97. doi: 10.1016/j.tplants.2009.11.009
- Szekowka, M., Heise, R., Tohge, T., Nunes-Nesi, A., Vosloh, D., and Huege, J., et al. (2013). Metabolic fluxes in an illuminated *Arabidopsis* Rosette. *Plant Cell* 25, 694–714. doi: 10.1105/tpc.112.106989
- Taji, T., Seki, M., Satou, M., Sakurai, T., Kobayashi, M., and Ishiyama, K., et al. (2004). Comparative genomics in salt tolerance between *Arabidopsis* and *Arabidopsis*-related halophyte salt cress using *Arabidopsis* microarray. *Plant Physiol.* 135, 1697–1709. doi: 10.1104/pp.104.039909
- Tarkowski, L. P., and Van den Ende, W. (2015). Cold tolerance triggered by soluble sugars: a multifaceted countermeasure. *Front. Plant Sci.* 6:203. doi: 10.3389/fpls.2015.00203
- Tauzin, A. S., and Giardina, T. (2014). Sucrose and invertases, a part of the plant defense response to the biotic stresses. *Front. Plant Sci.* 5:293. doi: 10.3389/fpls.2014.00293
- Thellier, M., and Lüttge, U. (2013). Plant memory: a tentative model. *Plant Biol.* 15, 1–12. doi: 10.1111/j.1438-8677.2012.00674.x
- Thibaud, M. C., Gineste, S., Nussaume, L., and Robaglia, C. (2004). Sucrose increases pathogenesis-related PR-2 gene expression in *Arabidopsis thaliana* through an SA-dependent but NPR1-independent signaling pathway. *Plant Physiol. Biochem.* 42, 81–8. doi: 10.1016/j.plaphy.2003.10.012
- Thomashow, M. F. (1999). Plant cold acclimation: freezing tolerance genes and regulatory mechanisms. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 50, 571–599. doi: 10.1146/annurev.arplant.50.1.571
- Tiburcio, A. F., Altabella, T., Bitrián, M., and Alcazar, R. (2014). The roles of polyamines during the lifespan of plants: from development to stress. *Planta* 240, 1–18. doi: 10.1007/s00425-014-2055-9
- Tognetti, J. A., Pontis, H. G., and Martínez-Noël, G. M. A. (2013). Sucrose signalling in plants—a world yet to be explored. *Plant Signal. Behav.* 8:e23316. doi: 10.4161/psb.23316
- Tohge, T., Schnell Ramos, M., Nunes-Nesi, A., Mutwil, M., Giavalisco, P., Steinhauser, D., et al. (2011). Toward the storage metabolome: profiling the Barley Vacuole. *Plant Physiol.* 157, 1469–1482. doi: 10.1104/pp.111.185710
- Urbanczyk-Wochniak, E., Baxter, C., Kolbe, A., Kopka, J., Sweetlove, L. J., and Fernie, A. R. (2005). Profiling of diurnal patterns of metabolite and transcript abundance in potato (*Solanum tuberosum*) leaves. *Planta* 221, 891–903. doi: 10.1007/s00425-005-1483-y
- Van Loon, L. C. (2007). Plant responses to plant growth-promoting rhizobacteria. *Eur. J. Plant Pathol.* 119, 243–254. doi: 10.1007/s10658-007-9165-1
- Vancova, R., Dobra, J., and Storchová, H. (2012). Recovery from drought stress in tobacco. *Plant Signal. Behav.* 7, 19–21. doi: 10.4161/psb.7.1.18375
- Van der Ent, S., Van Wees, S. C. M., and Pieterse, C. M. J. (2009). Jasmonate signaling in plant interactions with resistance-inducing beneficial microbes. *Phytochemistry* 70, 1581–1588. doi: 10.1016/j.phytochem.2009.06.009
- Van der Knaap, J. A., and Verrijzer, C. P. (2016). Undercover: gene control by metabolites and metabolic enzymes. *Genes Dev.* 30, 2345–2369. doi: 10.1101/gad.289140.116
- Vasilev, N., Boccard, J., Lang, G., Grömping, U., Fischer, R., Goepfert, S., et al. (2016). Structured plant metabolomics for the simultaneous exploration of multiple factors. *Sci. Rep.* 6:37390. doi: 10.1038/srep37390
- Wedeking, R., Maucourt, M., Deborde, C., Moing, A., Gibon, Y., Goldbach, H. E., et al. (2018). ¹H-NMR metabolomic profiling reveals a distinct metabolic recovery response in shoots and roots of temporarily drought-stressed sugar beets. *PLoS One* 13:e0196102. doi: 10.1371/journal.pone.0196102
- Weeda, S., Zhang, N., Zhao, X., Ndip, G., Guo, Y., Buck, G. A., et al. (2014). *Arabidopsis* transcriptome analysis reveals key roles of melatonin in plant defense systems. *PLoS One* 9:e93462. doi: 10.1371/journal.pone.0093462
- Xiao, W., Shen, J., and Jang, J. C. (2000). The role of hexokinase in plant sugar signal transduction and growth and development. *Plant Mol. Biol.* 44, 451–461. doi: 10.1023/A:1026501430422
- Xiao, Y., Savchenko, T., Baidoo, E. E. K., Chehab, W. E., Hayden, D. M., Tolstikov, V., et al. (2012). Retrograde signaling by the plastidial metabolite MECP regulates expression of nuclear stress-response genes. *Cell* 149, 1525–1535. doi: 10.1016/j.cell.2012.04.038

- Zeier, J. (2013). New insights into the regulation of plant immunity by amino acid metabolic pathways. *Plant Cell Environ.* 36, 2085–2103. doi: 10.1111/pce.12122
- Zhang, S. J., Yang, X., Sun, M. W., Sun, F., Deng, S., and Dong, H. S. (2009). Riboflavin-induced priming for pathogen defense in *Arabidopsis thaliana*. *J. Integr. Plant Biol.* 51, 167–174. doi: 10.1111/j.1744-7909.2008.00763.x
- Zhou, S., Lou, Y.-R., Tzin, V., and Jander, G. (2015). Alteration of plant primary metabolism in response to insect herbivory. *Plant Physiol.* 169, 1488–1498. doi: 10.1104/pp.15.01405
- Zimmerli, L., Métraux, J. P., and Mauch-Mani, B. (2001). Beta-aminobutyric acid-induced protection of *Arabidopsis* against the necrotrophic fungus *Botrytis cinerea*. *Plant Physiol.* 126, 517–523. doi: 10.1104/pp.126.2.517
- Zuther, E., Juszczak, I., Lee, Y. P., Baier, M., and Hinch, D. (2015). Time-dependent deacclimation after cold acclimation in *Arabidopsis thaliana* accessions. *Sci. Rep.* 5:12199. doi: 10.1038/srep12199
- Zuther, E., Schaarschmidt, S., Fischer, A., Erban, A., Pagter, M., Mubeen, U., et al. (2018). Molecular determinants of increased freezing tolerance due to low temperature memory in *Arabidopsis*. *Plant Cell Environ.* doi: 10.1111/pce.13502
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Schwachtje, Whitcomb, Firmino, Zuther, Hinch and Kopka. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Recruitment Model of Metabolic Evolution: Jasmonate-Responsive Transcription Factors and a Conceptual Model for the Evolution of Metabolic Pathways

Tsubasa Shoji*

Department of Biological Science, Nara Institute of Science and Technology (NAIST), Ikoma, Japan

OPEN ACCESS

Edited by:

Philipp Zerbe,
University of California, Davis,
United States

Reviewed by:

Justin William Walley,
Iowa State University, United States
Nobutaka Mitsuda,
National Institute of Advanced
Industrial Science and Technology
(AIST), Japan
Jacob Pollier,
Flanders Institute for Biotechnology,
Belgium

*Correspondence:

Tsubasa Shoji
t-shouji@bs.naist.jp

Specialty section:

This article was submitted to
Plant Metabolism and Chemodiversity,
a section of the journal
Frontiers in Plant Science

Received: 21 November 2018

Accepted: 12 April 2019

Published: 14 May 2019

Citation:

Shoji T (2019) The Recruitment Model
of Metabolic Evolution:
Jasmonate-Responsive Transcription
Factors and a Conceptual Model for
the Evolution of Metabolic Pathways.
Front. Plant Sci. 10:560.
doi: 10.3389/fpls.2019.00560

Plants produce a vast array of structurally diverse specialized metabolites with various biological activities, including medicinal alkaloids and terpenoids, from relatively simple precursors through a series of enzymatic steps. Massive metabolic flow through these pathways usually depends on the transcriptional coordination of a large set of metabolic, transport, and regulatory genes known as a regulon. The coexpression of genes involved in certain metabolic pathways in a wide range of developmental and environmental contexts has been investigated through transcriptomic analysis, which has been successfully exploited to mine the genes involved in various metabolic processes. Transcription factors are DNA-binding proteins that recognize relatively short sequences known as *cis*-regulatory elements residing in the promoter regions of target genes. Transcription factors have positive or negative effects on gene transcription mediated by RNA polymerase II. Evolutionarily conserved transcription factors of the APETALA2/ETHYLENE RESPONSE FACTOR (AP2/ERF) and basic helix-loop-helix (bHLH) families have been identified as jasmonate (JA)-responsive transcriptional regulators of unrelated specialized pathways in distinct plant lineages. Here, I review the current knowledge and propose a conceptual model for the evolution of metabolic pathways, termed “recruitment model of metabolic evolution.” According to this model, structural genes are repeatedly recruited into regulons under the control of conserved transcription factors through the generation of cognate *cis*-regulatory elements in the promoters of these genes. This leads to the adjustment of catalytic activities that improve metabolic flow through newly established passages.

Keywords: alkaloids, *cis*-regulatory element, jasmonates, recruitment model of metabolic evolution, regulon, specialized metabolism, terpenoids, transcription factor

INTRODUCTION

Biological processes generally depend on the coordinated expression of multiple genes. Transcription factors play a central role in controlling the RNA polymerase-mediated transcription of downstream genes. These genes form gene networks, or regulons, with transcription factors recognizing specific *cis*-regulatory elements in the promoter regions of these groups of target

genes. Complex, often long, metabolic pathways rely on the proper functioning of a large series of metabolic enzymes, membrane transporters, and other proteins. The activity of transcription factors often underlies the coordination of gene expression during various metabolic processes (Patra et al., 2013; Chezem and Clay, 2016; Zhou and Memelink, 2016).

A diverse range of specialized metabolites, such as bioactive alkaloids and terpenoids, are produced and accumulate in various plant species. These metabolites contribute to plant defense and reproduction in a changing environment. Due to their useful attributes, many natural products derived from plants or phytochemicals are utilized as medicines, drugs, dyes, perfumes, or other industrial materials. In contrast to universally present primary metabolites, the occurrence of specialized (or so-called secondary) metabolites is usually restricted to certain taxonomic groups. Metabolite levels are often highly variable, even within a single species or individual plant, reflecting the temporal and spatial dynamics of their production. Genomic and molecular approaches, often involving coexpression network analysis to select candidate genes (Yonekura-Sakakibara and Saito, 2013), have greatly facilitated the identification of structural genes involved in metabolic pathways. By contrast, the regulatory aspects of these genes, such as regulatory mechanisms at the transcriptional and other levels, have remained unexplored, representing a promising area of research in the coming years.

The central roles of MYB and basic helix-loop-helix (bHLH) family transcription factors in regulating the anthocyanin and related flavonoid pathways of many plant species have been a cornerstone example of the persistence of metabolic regulons comprising master transcriptional regulators with their downstream structural genes (Patra et al., 2013; Chezem and Clay, 2016). Well-studied instances include the regulation of the glucosinolate pathway by MYB family factors in *Arabidopsis* (Chezem and Clay, 2016). In notable contrast to the regulators that target a specific metabolic pathway (or set of related pathways), jasmonate (JA)-responsive factors in certain subgroups of the APETALA2/ETHYLENE RESPONSE FACTOR (AP2/ERF) and bHLH families have been found to be master regulators for a diverse range of specialized pathways, mostly for important alkaloids and terpenoids, in distinct plant lineages (Zhou and Memelink, 2016).

JAs are phytohormones derived from the octadecanoid pathway that play central roles as signaling molecules during biotic and abiotic stress responses in plants (Goossens et al., 2016a). Many specialized pathways are readily elicited by JA treatment (Goossens et al., 2016a; Zhou and Memelink, 2016). Indeed, many phytochemicals are thought to be involved in plant defense responses against pathogens and herbivores based on the JA-dependent elicitation of their biosynthetic pathways, along with their toxicity to biological agents. The perception of JA signals and the resulting cascades leading to gene regulation primarily occur via proteasome-dependent degradation of JAZ repressor proteins and the subsequent liberation of a few key transcription factors, including bHLH family member MYC2, from JAZ-mediated repression (Thines et al., 2007; Sheard et al., 2010; Zhang et al., 2015; Goossens et al., 2016a). It is important to address how the upstream JA signaling circuit is anchored to

downstream defense metabolism. A handful of the JA-responsive transcription factors of AP2/ERF and bHLH families, have been identified as missing links between the highly conserved JA signaling module and more divergent downstream pathways (Zhou and Memelink, 2016).

In this article, I provide an overview of the JA-responsive factors and their target metabolic pathways, which encompass a substantial portion of the specialized pathways for which transcriptional regulators have been defined (Patra et al., 2013; Zhou and Memelink, 2016). The identification of such evolutionarily conserved regulators targeting divergent pathways prompt me to contemplate how these metabolic regulons have been established during the evolution. This evolutionary issue is discussed and a conceptual model is proposed, mainly focusing on the JA-responsive factors and their regulons.

CLADE II, SUBGROUP IXa ERF TRANSCRIPTION FACTORS

Transcription factors of the AP2/ERF family are widespread in plants. The GCC-box (5'-AGCCGCC-3') element is a typical sequence recognized by AP2/ERF transcription factors. The DNA-binding AP2/ERF domain contains a three-stranded β -sheet followed by an α -helix, which form a unique interface required for DNA binding (Allen et al., 1998).

A group of AP2/ERF family transcription factors, including Octadecanoid-derivative Responsive Catharanthus AP2-domain (ORCA) proteins from *Catharanthus roseus* (Van der Fits and Memelink, 2000; Paul et al., 2017), OpERF2 from *Ophiorrhiza pumila* (Udomsom et al., 2016), ERF189 and ORC1 from tobacco (*Nicotiana tabacum*; Shoji et al., 2010; De Boer et al., 2011), JASMONATE RESPONSIVE ERF4 (JRE4)/GLYCOALKALOID METABOLISM9 (GAME9) from tomato and potato (Cárdenas et al., 2016; Thagun et al., 2016; Nakayasu et al., 2018), and AaORA from *Artemisia annua* (Lu et al., 2013), are classified into clade II of subgroup IXa (Nakano et al., 2006; Shoji et al., 2010, 2013). These transcription factors are involved in regulating JA-mediated defense metabolism in various plants. The JA-responsive ERF genes are present in a wide range of eudicots, usually as multicopy genes (Figure 1A). Multiple ERF genes are tandemly clustered on chromosomes in some plant genomes (Figure 1A). The phylogenetic relationships of ERFs from different species (Figure 1A) imply that independent generations of these gene clusters in distinct plant families through tandem gene duplication.

ORCAs in *C. roseus*

Terpenoid indole alkaloids (TIAs) are a large group of specialized products, including the valuable chemotherapy drugs vinblastine and vincristine. A variety of TIAs are derived from the key intermediate strictosidine, which is formed by condensation between tryptamine (a product of the shikimate pathway) and the seco-iridoid compound secologanin. TIA biosynthesis and its regulation have been intensively studied in the medicinally important species *C. roseus* (Apocynaceae) (Zhu et al., 2014). In *C. roseus*, ORCA2 and ORCA3 function as transcriptional

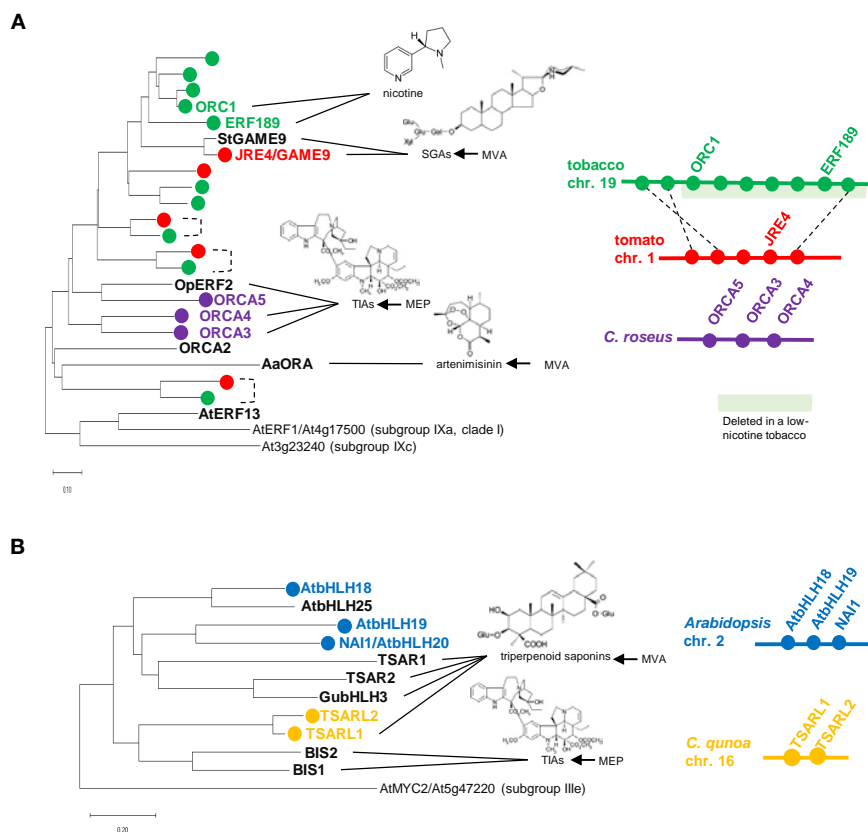


FIGURE 1 | Phylogenetic relationships and gene clustering of JA-responsive ERF (A) and bHLH (B) transcription factors of clade II, subgroup IXa and IVa, respectively. Full-length amino acid sequences were aligned using ClustalW (Thompson et al., 1994), and the trees were generated with the neighbor-joining algorithm using MEGA-X (Kumar et al., 2018). The *Arabidopsis* genes are included in the trees as an outgroup. The scale bars indicate the number of amino acid substitutions per residue. For isoprenoid-derived metabolites, the contribution of the mevalonate (MVA) or methylerythritol phosphate (MEP) pathway to their production is indicated on the right. Apparent orthologs from Solanaceae plants are connected by dotted lines. Gene clusters are depicted schematically. The region deleted in a low-nicotine tobacco cultivar (Kajikawa et al., 2017) is highlighted in light green.

regulators that induce the expression of TIA biosynthesis genes, including *strictosidine synthase* and *tryptophan decarboxylase*, encoding key enzymes in this pathway (Menke et al., 1999; Van der Fits and Memelink, 2000; Li et al., 2013). *ORCA3* is physically linked to *ORCA4* and *ORCA5*, forming a gene cluster in the genome (Figure 1A, Paul et al., 2017). *ORCA4* shares an overlapping function with *ORCA2* and *ORCA3*, but *ORCA4* also targets additional TIA genes (Paul et al., 2017). Unlike *ORCA2* (Li et al., 2013) and *ORCA3* (Van der Fits and Memelink, 2000), the overexpression of *ORCA4* causes a drastic increase in TIA accumulation (Paul et al., 2017). *C. roseus* MYC2 (CrMYC2) directly upregulates the expression of *ORCA3* by recognizing a G-box-like element in its promoter (Zhang et al., 2011), and it also coregulates TIA structural genes with *ORCA3* (Paul et al., 2017). In addition to their role in transcriptional regulation, ORCs and CrMYC2 are activated by phosphorylation by a kinase involved in a JA-activated MAP kinase cascade (Paul et al., 2017).

OpERF2 in *Ophiorrhiza pumila*

Camptothecin is an antitumor TIA that functions by inhibiting DNA topoisomerase I activity. This clinically important TIA

is produced by various angiosperms from taxonomically distant families, including *Ophiorrhiza pumila* (Rubiaceae) (Sirikantaramas et al., 2007). *OpERF2* was originally isolated from *O. pumila* hairy roots. The suppressed expression of this gene resulted in the reduced expression of genes involved in seco-iridoid and upstream methylerythritol phosphate (MEP) pathways, which supply secologanin for downstream camptothecin production, although this did not have a significant impact on TIA accumulation (Udomsorn et al., 2016).

ERF189 and ORC1 in Tobacco

Nicotine is composed with two heterocyclic rings, the ornithine-derived pyrrolidine ring and the nicotinate-derived pyridine ring. In tobacco (*Nicotiana tabacum*, Solanaceae), this toxic alkaloid is produced in roots and primarily accumulates in leaves, functioning as a defense compound against herbivores (Dewey and Xie, 2013). Tobacco *ERF189*, *ORC1*, and related *ERF* genes are clustered together in the tobacco genome (Figure 1A, Shoji et al., 2010; De Boer et al., 2011; Kajikawa et al., 2017). A cluster of *ERFs* including *ERF189* and *ORC1* were found to be deleted to a large extent in a tobacco cultivar with low

nicotine content (**Figure 1A**, Shoji et al., 2010; Kajikawa et al., 2017). Although not yet proven, ERF189 is considered to be a primary transcriptional regulator of nicotine biosynthesis, given that its expression profiles are similar to those of the downstream biosynthesis genes: strong expression in roots (Kajikawa et al., 2017), no induction in response to NaCl (Shoji and Hashimoto, 2015; Kajikawa et al., 2017), and the suppression of JA-dependent induction by ethylene (Shoji et al., 2000, 2010). A large series of metabolic and transport genes in this pathway are upregulated by ERF189, which recognizes P-box elements, but not the typical GCC-box elements, in their promoters (Shoji et al., 2010, 2013; Shoji and Hashimoto, 2011a). Tobacco MYC2 regulates the expression of *ERF189* and directly activates the transcription of nicotine biosynthesis genes together with ERF189 (Shoji and Hashimoto, 2011b; Zhang et al., 2012).

JRE4/GAME9 in Tomato and Potato

Steroidal glycoalkaloids (SGAs) are cholesterol-derived, nitrogen-containing metabolites found in the inedible parts of Solanaceae plants such as tomato (*Solanum lycopersicum*) and potato (*S. tuberosum*) (Cárdenas et al., 2015). In the tomato and potato genomes, the *JRE4/GAME9* gene is present in a cluster with related *ERF* genes (Cárdenas et al., 2016; Thagun et al., 2016). *JRE4/GAME9* regulates nearly an entire series of SGA metabolic steps, including those in the upstream isoprenoid-producing mevalonate (MVA) pathway (Cárdenas et al., 2016; Thagun et al., 2016; Nakayasu et al., 2018). A loss of *JRE4/GAME9* function drastically reduced SGA accumulation and resistance to chewing insects in tomato, demonstrating the major role of this transcription factor in defense-related SGA formation (Nakayasu et al., 2018). Tomato MYC2 and *JRE4/GAME9* synergistically activated the promoters of SGA genes in tobacco protoplasts (Cárdenas et al., 2016). In agreement with the results of promoter binding studies, cognate *cis*-regulatory elements are significantly enriched in the proximal promoter regions of SGA biosynthesis genes, supporting the direct regulation of these genes by *JRE4/GAME9* (Thagun et al., 2016). A comparison of the genomes of ancestral and cultivated species of the *Solanum* genus pointed to the possible selection of certain alleles of *JRE4/GAME9* during domestication, which might have contributed to the decrease in antinutritional SGA levels in cultivated *Solanum* species (Hardigan et al., 2017; Zhu et al., 2018).

AaORA in *Artemisia annua*

Artemisinin, a sesquiterpene lactone produced by the traditional Chinese herb *Artemisia annua* (Asteraceae), has been exploited as an effective anti-malaria agent (Tang et al., 2014). *A. annua* ORA (AaORA) is a transcriptional regulator of artemisinin biosynthesis that upregulates the expression of genes involved in this pathway, including *amorpha-4,11-diene synthase*, *CYP71AV1*, and *double bond reductase 2* (Lu et al., 2013). *AaORA* is specifically expressed in the trichomes of aerial organs, including artemisinin-producing glandular trichomes (Olofsson et al., 2011; Lu et al., 2013). Since numerous transcription factors from various families (e.g., bHLH, ERF, bZIP, and WRKY), in addition to AaORA, were shown to regulate artemisinin

biosynthesis (Tang et al., 2014; Lv et al., 2017), the relative importance of each transcription factor and their functional relationships in this process should be addressed.

Arabidopsis thaliana ERF13

AtERF13 is the only member of clade II, subgroup IXa in *Arabidopsis thaliana* (Brassicaceae). In contrast to the ERFs mentioned above, *AtERF13* was not yet shown to be involved in a specific metabolic pathway. *AtERF13* is induced in response to a range of biotic and abiotic stresses, such as JA, wounding, insect feeding, colonization of beneficial bacteria, high osmolality, and NaCl (Lee et al., 2010; Sogabe et al., 2011; Srivastava et al., 2012; Schweizer et al., 2013). *AtERF13* binds to COUPLING ELEMENT1 (CE1), a *cis*-regulatory element required for abscisic acid (ABA)-responsive gene expression, and the overexpression of *AtERF13* confers increased sensitivity to ABA in *Arabidopsis*, suggesting this gene functions in abiotic stress resistance (Lee et al., 2010). *AtERF13* is also involved in resistance to insect herbivores, acting downstream of MYC2 (a central player in JA signaling) and mediating the expression of a subset of MYC2-regulated defense genes (Schweizer et al., 2013). *AtERF13* is phosphorylated at its tyrosine residues, as revealed by phosphoproteomic analysis, suggesting that its activity is regulated via post-translational modification (Nemoto et al., 2015).

SUBGROUP IVa bHLH TRANSCRIPTION FACTORS

Another group of JA-responsive transcription factors is attracting attention as regulators of metabolic pathways in diverse plants. These transcription factors include bHLH IRIDOID SYNTHESIS1 (BIS1) and BIS2 from *C. roseus* (Van Moerkercke et al., 2015, 2016), TRITERPENE SAPONIN BIOSYNTHESIS ACTIVATING REGULATOR1 (TSAR1) and TSAR2 from *Medicago truncatula* (Mertens et al., 2016a), TSAR-LIKE1 (TSARL1) from *Chenopodium quinoa* (Jarvis et al., 2017), GubHLH3 from *Glycyrrhiza uralensis* (Tamura et al., 2018), and bHLH18, bHLH19, bHLH20/NAI1, and bHLH25 from *Arabidopsis* (Matsushima et al., 2004), which all belong to subgroup IVa of the bHLH family (**Figure 1B**, Heim et al., 2003; Goossens et al., 2016b).

Unlike AP2/ERF family members, which are specific to plants, the bHLH transcription factor family is widely present in eukaryotic organisms and has expanded, especially in land plants (Feller et al., 2011). The signature bHLH domain is composed of an N-terminal basic region that binds to negatively charged DNA and a helix-loop-helix motif responsible for protein dimerization. bHLH transcription factors form homo- or heterodimers that typically bind to E-box (5'-CANNTG-3') elements, such as G-box (5'-CACGTG-3') and N-box (5'-CACGAG-3') elements, in the promoter regions of their target genes.

BISs in *C. roseus*

In addition to ORCAs and CrMYC2, BIS1 and BIS2, a pair of homologous JA-responsive bHLH transcription factors, are involved in regulating TIA formation in *C. roseus*. BISs

specifically act on a branch of the TIA pathway that supplies the seco-iridoid intermediate, secologanin, for incorporation into TIAs. Overexpression of *BIS1* or *BIS2* results in the upregulation of genes involved in seco-iridoid and upstream MEP pathways, thereby increasing the accumulation of downstream TIAs (Van Moerkercke et al., 2015, 2016). The finding that *BIS2* is induced by *BIS1* or *BIS2* overexpression points to the existence of a positive feedback loop (Van Moerkercke et al., 2016). In contrast to subgroup IIIe MYC2 transcription factors, BISs cannot interact with JAZ proteins and are thus not direct targets of the repressors integrated into the JA signaling module (Van Moerkercke et al., 2016).

TSARs in *Medicago truncatula*

The model legume plant *Medicago truncatula* (Fabaceae) produces oleanane-type triterpenoid saponins. These amphipathic glycosides, containing triterpenoid aglycones, exhibit a diverse range of biological activities (Osborn et al., 2011). In *M. truncatula*, TSAR1 and TSAR2, two homologous bHLH transcription factors, are JA-responsive transcriptional regulators of triterpenoid saponin biosynthesis (Mertens et al., 2016a). While the isoprenoid-producing MVA pathway is commonly targeted by both TSARs, TSAR1 and TSAR2 specifically regulate two distinct downstream branches of this pathway, producing nonhemolytic and hemolytic saponins, respectively (Mertens et al., 2016a). TSARs activate the gene encoding 3-Hydroxy-3-Methylglutaryl-CoA Reductase, a rate-limiting enzyme in the MVA pathway, by directly recognizing the N-box element in its promoter (Mertens et al., 2016a).

TSARL1 in *Chenopodium quinoa*

Chenopodium quinoa (Chenopodiaceae), or quinoa, is a staple food crop in Andean countries. Quinoa seeds have high nutritional value, but bitterness of the seeds due to the accumulation of triperpenoid saponins (oleanane-type) is disadvantageous (Kuljanabhagavad et al., 2008). In *C. quinoa*, *TSARL1* and *TSARL2* are clustered together (Figure 1B) and are expressed in seeds and roots, respectively (Jarvis et al., 2017). In sweet quinoa strains, loss-of-function mutations of *TSARL1*, including one that appears to cause alternative splicing, allowed the down-regulation of genes involved in the production of the antinutritional saponins (Jarvis et al., 2017).

GubHLH3 in *Glycyrrhiza uralensis*

The medicinal legume *Glycyrrhiza uralensis* (Fabaceae) is rich in oleanane-type triterpenoid saponins, such as glycyrrhizin, which is used as a pharmaceutical compound and sweetener, as well as soyasaponins (Hayashi and Sudo, 2009). *G. uralensis* bHLH3 (GubHLH3), a JA-responsive bHLH transcription factor, upregulates the expression of soyasaponin biosynthesis genes, such as those encoding CYP93E3 and CYP72A566, which are involved in oxidative modifications of the triterpenoid backbone (Tamura et al., 2018). Consistently, the overexpression of *GubHLH3* increased the levels of soyasapogenol B and other intermediates of the soyasaponin pathway in *G. uralensis* hairy roots (Tamura et al., 2018).

bHLH18, bHLH19, bHLH20/NAI1, and bHLH25 in *Arabidopsis*

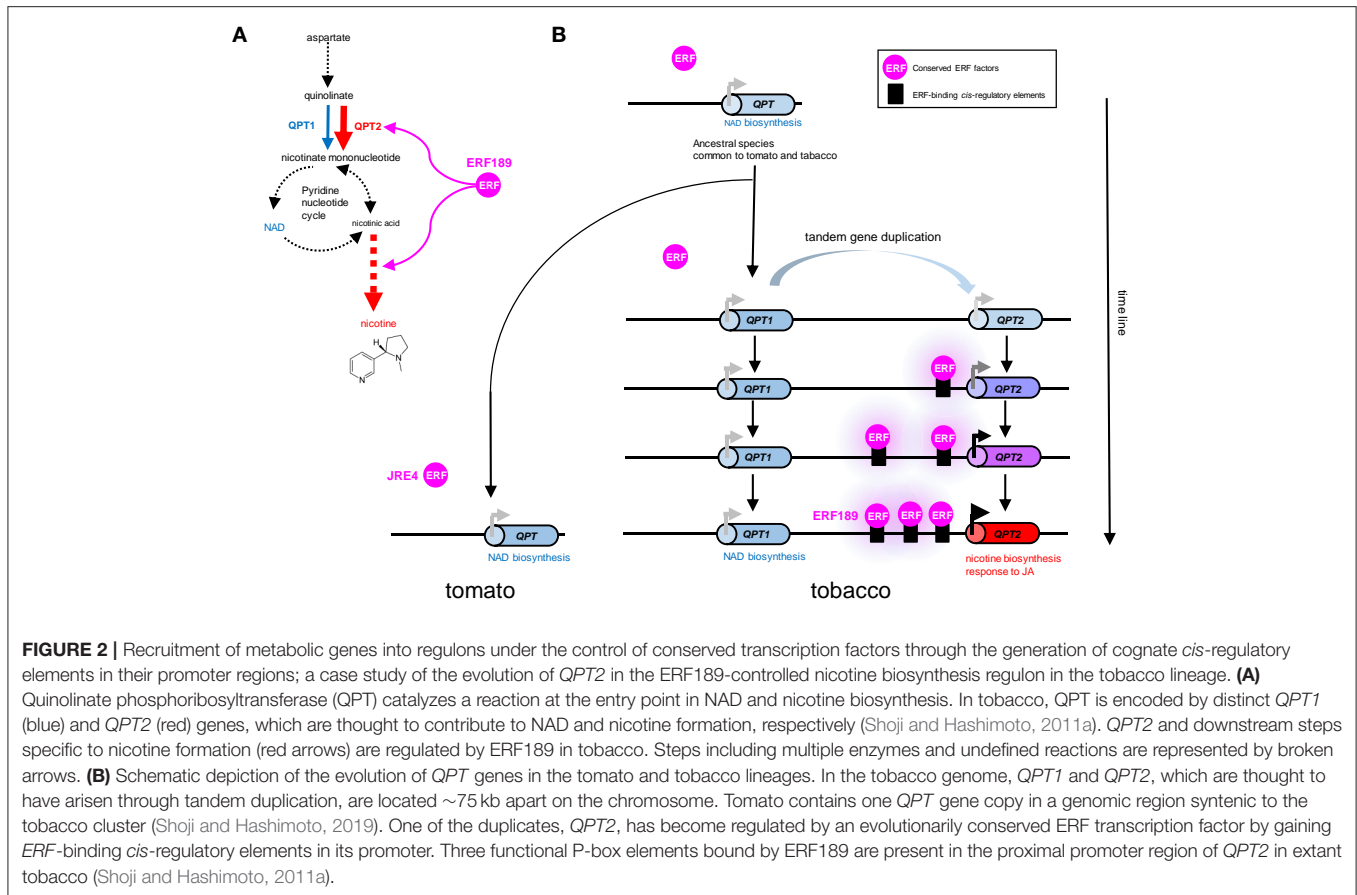
In *Arabidopsis*, four genes, *bHLH18*, *bHLH19*, *bHLH20/NAI1*, and *bHLH25*, encode subgroup IVa bHLH transcription factors; three of them, except *bHLH25*, form a gene cluster (Figure 1B). *NAI1*, which resides in this three-gene cluster, is indispensable for the formation of the ER body, an ER-derived rod-shape organelle found in plants of the *Brassicales* order (Matsushima et al., 2004). ER bodies are constitutively present in *Arabidopsis* seedlings and roots. By contrast, in rosette leaves, wounding and JA treatment induce the formation of this defense-related organelle, which accumulates large amounts of β -glucosidases, whose activities increase when the compartment is disrupted (Nakano et al., 2014). *NAI1* regulates the expression of genes encoding proteins required for ER body formation and activity, including *PYK10*, a major β -glucosidase in this organelle. *PYK10* functions as a myrosinase that hydrolyzes indole glucosinolates, a group of important defense compounds in *Arabidopsis* and related species (Nakano et al., 2017). The phylogenetic co-occurrence of ER bodies and indole glucosinolates and the co-expression of the associated genes also support the functional coordination between this organelle and glucosinolate metabolism (Nakano et al., 2017).

A previous study suggested the involvement of bHLH18, bHLH19, bHLH20/NAI1, and bHLH25 in JA-mediated inhibition of iron uptake in *Arabidopsis* (Cui et al., 2018). JA represses iron uptake by promoting the degradation of FIT/bHLH29, a central transcriptional regulator of iron-uptake genes critical to metal homeostasis (Cui et al., 2018). The four subgroup IVa bHLHs interact with FIT protein and promote its JA-stimulated removal through proteasome-dependent degradation (Cui et al., 2018).

THE GAIN OF *cis*-REGULATORY ELEMENTS

The recruitment of metabolic genes into regulons likely requires the gain of transcription factor-binding *cis*-regulatory elements in the appropriate promoter regions. Such a process is fairly likely, considering the relatively frequent, simple generation of short sequence elements in noncoding promoter regions that can have a degree of redundancy and acquire functions through mutational changes, such as point mutations and transpositions (Wray, 2007; Swinnen et al., 2016).

A case study of a tobacco gene involved in nicotine biosynthesis supports such a scenario. Quinolinate phosphoribosyltransferase (QPT) is a primary metabolic enzyme involved in NAD biosynthesis in all organisms. However, in tobacco, QPT also supplies a significant amount of intermediates required for downstream nicotine biosynthesis (Figure 2A). To satisfy such a metabolic demand, tandem duplication of *QPT* has occurred in the *Nicotiana* lineage, generating a cluster of *QPT1* and *QPT2* genes (Figure 2, Shoji and Hashimoto, 2019). These genes are thought to be involved in NAD and nicotine biosynthesis, respectively, based on their distinct expression



patterns (Shoji and Hashimoto, 2011a). *QPT2* harbors multiple ERF189-binding P-box elements in its promoter required for its transcriptional activation by ERF189 (Shoji and Hashimoto, 2011a). The progressive acquisition of these elements after gene duplication has ensured the involvement of *QPT2* in the ERF189-controlled regulon of the nicotine pathway (Figure 2B). The frequent occurrence of JRE4/GAME9 binding elements in the proximal promoter regions of SGA biosynthesis genes in tomato implies that such a notion is also applicable to these genes (Thagun et al., 2016).

EVOLUTIONARY CHANGES IN TRANSCRIPTION FACTORS

In contrast to the gains (and losses) of *cis*-regulatory elements that strongly contribute to the rewiring of gene regulatory networks, mutational changes in transcription factors, which have profound, pleiotropic effects on numerous downstream genes, are relatively constrained. Nevertheless, there are also examples of the modification of the functionalities and expression patterns of *trans*-acting factors (Maerkl and Quake, 2009).

A series of subgroup IXa ERFs have divergent DNA-binding specificities to GCC-box elements and to related but distinct P-box and CS1-box elements. Such distinct binding specificities

can be accounted for by a few amino acid differences in a small stretch of the DNA-binding domain (Figure 3, Shoji et al., 2013). It appears that a progressive evolutionary trajectory has led from transcription factors that recognize only a canonical GCC-box to *Nicotiana*-specific ERF189-type transcription factors that bind to P-box but not GCC-box elements via functional intermediates, such as ORCA3-type transcription factors, which bind to multiple elements, including both GCC-box and P-box elements (Figure 3, Shoji et al., 2013). The development of unique combinations of *cis*-elements and *trans*-factors may have been indispensable for avoiding missed connections among unrelated regulatory circuits and, thus, the establishment of lineage-specific specialized pathways. This process appears to have occurred independently of the development of a broad range of ERFs targeting GCC-box elements involved in general defense responses.

Nicotine and SGA biosynthesis pathways in distinct lineages of the same Solanaceae family, which are regulated by orthologous ERFs, share many properties, such as JA-dependent induction and suppression by ethylene (Shoji et al., 2010; Nakayasu et al., 2018). By contrast, the site of their biosynthesis differs between the two lineages: nicotine is synthesized exclusively in tobacco roots, whereas SGAs are produced in nearly all inedible parts of tomato and potato, including leaves and roots. This difference depends on the differential expression patterns of the transcriptional regulators ERF189

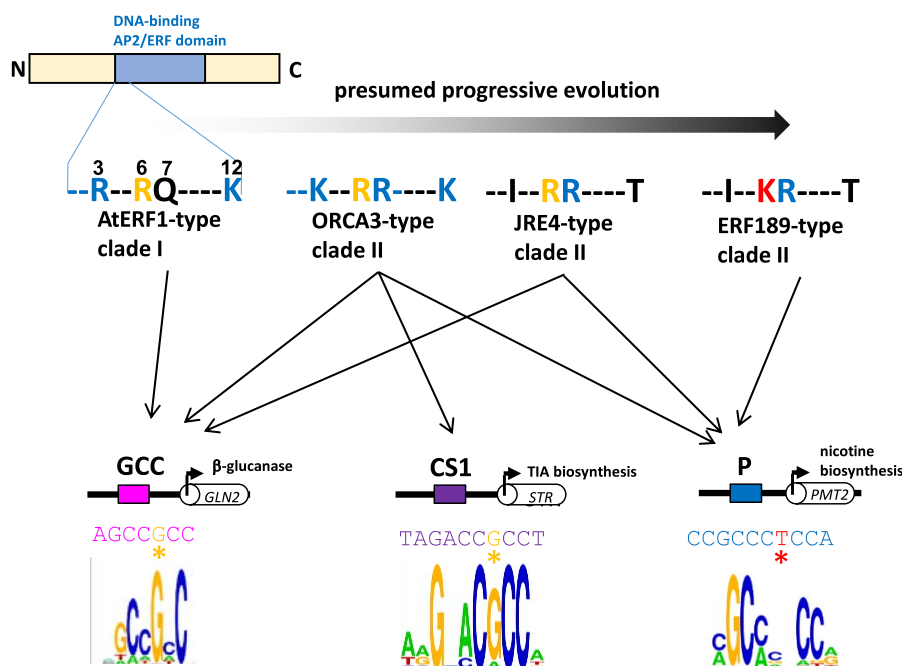


FIGURE 3 | Divergent DNA-binding specificities of subgroup IXa ERF transcription factors. Amino acid residues at positions 3, 6, 7, and 12 in an N-terminal portion of the DNA-binding AP2/ERF domain are shown (counting from the end of the domain). Basic residues R and K at positions 3, 7, and 12 (blue) increase the nonspecific affinity of the proteins for DNA by interacting with negatively charged phosphate backbones (Shoji et al., 2013). The substitution of one residue at position 6 from R (yellow) to K (red) in ERF189-type transcription factors alters the recognition of nucleotide bases (G to T) at a specific position (marked with asterisks) in the *cis*-regulatory elements (Shoji et al., 2013). Perhaps progressive evolution has occurred from ancestral AtERF1-type transcription factors to more specialized ERF189-type transcription factors via functional intermediates with amino acid substitutions at a few positions. The *in vitro* binding specificities of recombinant ORCA3 are represented by sequence logos (Shoji et al., 2013) GLN2, β-1,3-glucanase; STR, strictosidine synthase; PMT2, putrescine *N*-methyltransferase.

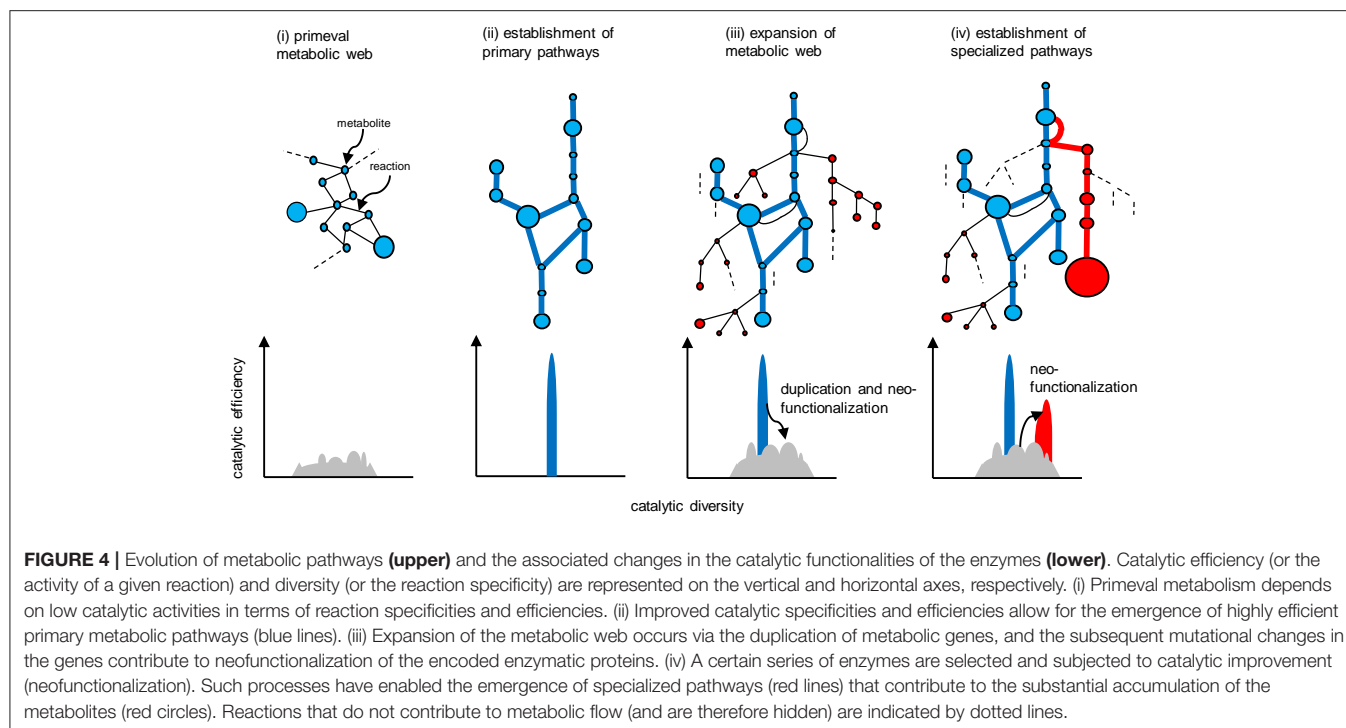
and JRE4/GAME9 (Cárdenas et al., 2016; Thagun et al., 2016; Kajikawa et al., 2017). To guarantee the function of each group of metabolites in plant defense, the tissue-specific expression patterns of these regulators may have developed independently after the separation of the two lineages, whereas their responses to JA and other features have been conserved between lineages. These ideas point to the elastic evolution of sets of a particular transcription factor and its downstream metabolic genes as independent units with specialized roles in chemical defense in a lineage-specific manner.

Despite the functional differences noted above, the JA-responsive transcription factors are considered components of conserved regulatory mechanisms present in various species. For instance, in transgenic tomato plants, a promoter reporter of tobacco *QPT2* regulated by ERF189 was expressed in a JA-responsive and cell type-specific manner, as in tobacco, and this expression was mediated by JRE4 (Shoji and Hashimoto, 2019). TSARs and BISs were shown to be functionally exchangeable, as the orthologous bHLH factors regulate each other's target genes in *C. roseus* and *M. truncatula* (Mertens et al., 2016b). Both of these examples clearly point to the interchangeable nature of these factors among species with entirely different pathways (e.g., ornithine-derived nicotine vs. MVA pathway-derived SGAs for ERF189 and JRE4, and MVA pathway-derived saponins vs. MEP pathway-derived TIAs for TSARs and BISs), supporting the functional conservation of these factors.

RECRUITMENT MODEL OF METABOLIC EVOLUTION

Metabolism is a fundamental requirement of all living organisms. Primeval metabolism, or the simple conversion of substances, is thought to rely on a small number of proteinaceous or other catalysts with low reaction specificities and efficiencies (Figure 4i). Metabolic systems have evolved toward increasing order and efficiency (Weng et al., 2012). Contemporary primary metabolism, which was established early and has been maintained, is carried out by robust systems mediated by enzymes with high specificities and efficiencies (Figure 4ii).

Enzymes involved in specialized metabolism are thought to have emerged through duplication, beginning with sophisticated primary enzymes as progenitors, followed by mutational changes in the duplicates and leading to neofunctionalization of the enzymes (Figure 4iii). Relaxed constraints on the specificities and efficiencies of newly generated duplicates allow them to explore a wide range of catalytic possibilities. The promiscuity of multifunctional enzymes, with broader specificities emerging through this process, contributes to the expansion of the metabolic web (Figure 4iii). This web even includes the virtual activities of hidden enzymes (dotted lines in Figure 4) that do not contribute to actual metabolic flow due to limited substrate availability or marginal enzymatic activity; such hidden activities are not readily eliminated by (and are more tolerant to) selection.



If these changes are not deleterious or neutral and are thus not eliminated by purifying selection, the metabolic grids continue to build through neutral evolutionary processes such as genetic drift. It seems reasonable that autotrophic plants accumulate low-molecular-weight metabolites derived from photosynthetic assimilates, which usually have antioxidant properties to some extent and are often sequestered in cellular compartments such as vacuoles. These natural products, including those that accumulate in trace amounts, do not necessarily have adaptive significance (Koonin, 2016).

The emergence of specialized pathways allowing for the efficient production and accumulation of substantial amounts of metabolites requires the selection of specific flows from the expanded metabolic web, again increasing order and efficiency (**Figure 4iv**). This process largely relies on positive natural selection rather than neutral evolution, which is dependent on randomness. I propose a conceptual model, recruitment model of metabolic evolution, describing this process. According to this model, structural genes are repeatedly recruited into regulons under the control of evolutionarily conserved transcription factors (which should be activators rather than repressors), such as the JA-responsive ERFs and bHLHs (**Figure 5**). When a gene in the metabolic web becomes regulated by a transcription factor, obtaining cognate *cis*-regulatory elements, metabolic flows are generated or altered accordingly. Although such events readily occur at high frequency, most of these mutational changes are immediately eliminated and are not maintained in the population. On the other hand, when the newly generated flows result in the accumulation of beneficial products, such as defense compounds, conferring adaptive advantages to the plant, the probability that such changes will be maintained and eventually fixed in the population increases tremendously. Once

the beneficial flows occur, the likelihood that mutational events that enhance these flows (such as the transcriptional activation of other metabolic genes and the optimization of catalytic specificities and efficiencies associated with the flows) is expected to rise markedly as well. An initial, mostly accidental, event creating new metabolic flows may trigger cascading mutational changes associated with and improving the flows, eventually leading to the establishment of metabolic regulons and pathways, perhaps within a relatively short evolutionary timescale. The bioinformatic and mathematical bases of the model remain to be explored.

The extensive rewiring of transcriptional circuits alters metabolic regulons that were once established under the original transcription factors (**Figure 6**). During such processes, takeover of the regulons by new transcription factors (including those derived from the original transcription factors by duplication) and the associated changes in the connections in the circuits occur frequently through changes in *cis*-regulatory elements and transcription factors (**Figure 6**, Johnson, 2017). Contemporary regulatory networks are often complex and include multiple transcription factors, which act as either activators or repressors, and in some cases regulate only specific parts of pathways (e.g., ORCA3 regulates some but not all TIA genes). Extensive rewiring of circuits may contribute to the advent of the complicated regulatory organization found in extant metabolic pathways, which also could account for the fact that few metabolic pathways have simple regulons comprising only a single master regulator and its downstream structural genes.

Analyses of metabolic evolution have emphasized mutational changes to catalytic enzymes (Weng et al., 2012; Moghe and Last, 2015). If metabolic genes are

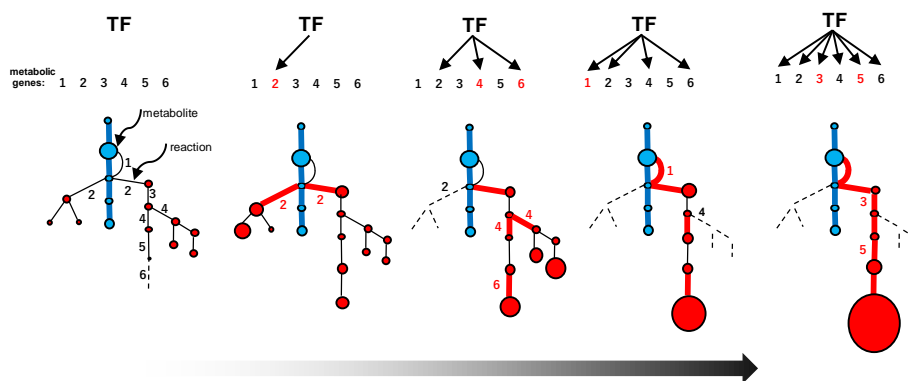


FIGURE 5 | Recruitment model of metabolic evolution. The emergence of an efficient specialized pathway (red lines) from an expanded metabolic web, including the promiscuous steps (part of the web shown in **Figure 4iii**), is depicted schematically. Metabolic genes and the corresponding steps are numbered (1–6). The recruitment of metabolic genes and enzymes into the regulon under the control of a conserved transcription factor (TF) occurred repeatedly (red numbers). The catalytic specificity of promiscuous enzymes 2 and 4 was optimized (black numbers in the web). The recruitment of genes affecting metabolic flows in the web (which increase the possibility of recruiting other genes into the regulon) and catalytic optimization further enhance favorable flows once established.

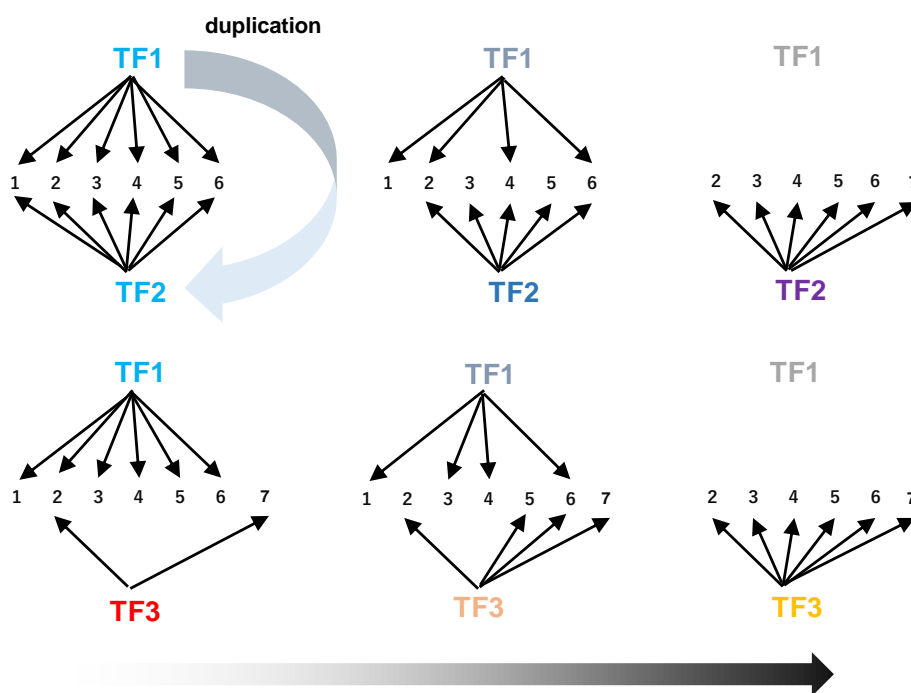


FIGURE 6 | Rewiring of transcriptional circuits. An original regulon (including genes 1 to 6) controlled by TF1 is hijacked by other transcription factors (TF2 and TF3) and eventually modified (removal of 1 and addition of 7). TF2 arose from TF1 through gene duplication. The colors of the TFs reflect functional changes in these proteins during evolution.

not functionally expressed and no flows are associated with these genes, they are not subjected to positive or purifying selection. The recruitment model, which presumes that transcription factors activate genes prior to changes in catalytic activity, adequately addresses this point as well.

Metabolic evolution is driven by functional changes in catalytic enzymes and changes in the expression patterns of metabolic genes. There appear to be limits to the changes

made to the catalytic functionalities of enzymes belonging to a limited number of protein families without hampering the structural and functional stability of protein frames. Therefore, changes in the combinations of metabolic genes with specific temporal and spatial expression patterns might also have significantly contributed to the rise in chemodiversity found in plants. Plants have exploited their limited repertoire of enzymes in a combinatorial manner to produce these diverse compounds.

PERSPECTIVES OF “EVO-META” BIOLOGY

“Nothing in biology makes sense except in the light of evolution” is a famous quotation by Dr. Theodosius Dobzhansky, a Ukrainian-American evolutionary geneticist (Dobzhansky, 1973). The discovery of homeotic genes encoding a group of transcription factors that direct the organization of the body plans of vertebrates and invertebrates has helped elucidate the evolutionarily conserved mechanisms governing development, leading to the rise of Evolutionary Developmental (Evo-Devo) biology. Classic anatomy and embryology, as well as modern developmental biology, share some affinity with evolutionary biology. Paleontology based on fossil evidence is one of the main areas of focus in evolutionary biology.

Chemodiversity of specialized products in plants has been shaped through (and is a product of) evolution. Unfortunately, it is difficult to predict the hues and fragrances of ancient flowers from extinct plants. Nevertheless, the chemodiversity found in extant species and the diverse series of plant genomes, including those yet to be explored, is highly informative (Nakamura et al., 2014). A long period of collection of natural products with divergent chemical structures and biological activities, along with a better understanding of the grouping of biosynthetic pathways and associated enzymes, has led to an awareness of some sort of order behind this chemodiversity. Elucidating the molecular biology behind regulatory factors, such as master

transcription factors, that orchestrate these metabolic processes is expected to reveal the universal principles ruling the metabolic processes that produce a diverse range of specialized products. The challenges of “Evo-Meta” (Evolutionary Metabolic) biology aimed at uncovering the origins of this chemodiversity are just beginning.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

This work was supported in part by the Japan Society for the Promotion of Science (Grants-in-Aids for Scientific Research number 17K07447 to TS).

ACKNOWLEDGMENTS

I thank Drs. Takashi Hashimoto and Yasuyuki Yamada (NAIST) for their long-term collaboration and continuous encouragement. I appreciate the encouraging and constructive remarks by Dr. Shigetada Nakanishi (SUNBOR) on our research related to this article.

REFERENCES

- Allen, M. D., Yamasaki, K., Ohme-Takagi, M., Tateno, M., and Suzuki, M. (1998). A novel mode of DNA recognition by a β -sheet revealed by the solution structure of the GCC-box binding domain in complex with DNA. *EMBO J.* 18, 5484–5496. doi: 10.1093/emboj/17.18.5484
- Cárdenas, P. D., Sonawane, P. D., Heinig, U., Bocobza, S. E., Burdman, S., and Aharoni, A. (2015). The bitter side of the nightshades: genomics drives discovery in Solanaceae steroidal alkaloid metabolism. *Phytochemistry* 113, 24–32. doi: 10.1016/j.phytochem.2014.12.010
- Cárdenas, P. D., Sonawane, P. D., Pollier, J., Bossche, R. V., Dewangan, V., Weithorn, E., et al. (2016). GAME9 regulates the biosynthesis of steroidal alkaloids and upstream isoprenoids in the plant mevalonate pathway. *Nat. Commun.* 7:10657. doi: 10.1038/ncomms10654
- Chezem, W. R., and Clay, N. K. (2016). Regulation of plant secondary metabolism and associated specialized cell development by MYBs and bHLHs. *Phytochemistry* 131, 26–43. doi: 10.1016/j.phytochem.2016.08.006
- Cui, Y., Chen, C. L., Cui, M., Zhou, W. J., Wu, H. L., and Ling, H. Q. (2018). Four IVa bHLH transcription factors are novel interactors of FIT and mediate JA inhibition of iron uptake in Arabidopsis. *Mol. Plant* 11, 1166–1183. doi: 10.1016/j.molp.2018.06.005
- De Boer, K., Tillemann, S., Pauwels, L., Vanden Bossches, R., De Sutter, V., Vanderhaeghen, R., et al. (2011). APETALA2/ETHYLENE RESPONSE FACTOR and basic helix-loop-helix tobacco transcription factors cooperatively mediate jasmonate-elicited nicotine biosynthesis. *Plant J.* 66, 1053–1065. doi: 10.1111/j.1365-3113X.2011.04566.x
- Dewey, R. E., and Xie, J. (2013). Molecular genetics of alkaloid biosynthesis in *Nicotiana tabacum*. *Phytochemistry* 94, 10–27. doi: 10.1016/j.phytochem.2013.06.002
- Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *Am. Biol. Teach.* 35, 125–129. doi: 10.2307/4444260
- Feller, A., Machemer, K., Braun, E. L., and Grotewold, E. (2011). Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *Plant J.* 66, 94–116. doi: 10.1111/j.1365-3113X.2010.04459.x
- Goossens, J., Fernández-Calvo, P., Schweizer, F., and Goossens, A. (2016a). Jasmonates: signal transduction components and their roles in environmental stress responses. *Plant Mol. Biol.* 91, 673–689. doi: 10.1007/s11103-016-0480-9
- Goossens, J., Mertens, J., and Goossens, A. (2016b). Role and functioning of bHLH transcription factors in jasmonate signaling. *J. Exp. Bot.* 68, 1333–1347. doi: 10.1093/jxb/erw440
- Hardigan, M. A., Laimbeer, F. P. E., Newton, L., Crisovan, E., Hamilton, J. P., Vaillancourt, B., et al. (2017). Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato. *Proc. Natl. Acad. Sci. U.S.A.* 114, E9999–E10008. doi: 10.1073/pnas.1714380114
- Hayashi, H., and Sudo, H. (2009). Economic importance of licorice. *Plant Biotechnol.* 26, 101–104. doi: 10.5511/plantbiotechnology.26.101
- Heim, M. A., Jakoby, M., Werber, M., Martin, C., Weisshaar, B., and Bailey, P. C. (2003). The basic helix-loop-helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity. *Mol. Biol. Evol.* 20, 735–747. doi: 10.1093/molbev/msg088
- Jarvis, D. E., Ho, Y. S., Lightfoot, D. J., Schöckel, S. M., Li, B., Borm, T. J. A., et al. (2017). The genome of *Chenopodium quinoa*. *Nature* 542, 307–312. doi: 10.1038/nature21370
- Johnson, A. D. (2017). The rewiring of transcription circuits in evolution. *Curr. Opin. Genet. Dev.* 47, 121–127. doi: 10.1016/j.gde.2017.09.004
- Kajikawa, M., Sierro, N., Kawaguchi, H., Bakaher, N., Ivanov, N. V., Hashimoto, T., et al. (2017). Genomic insights into the evolution of the nicotine biosynthesis pathway in tobacco. *Plant Physiol.* 174, 999–1011. doi: 10.1104/pp.17.00070
- Koonin, E. V. (2016). Splendor and misery of adaptation, or the importance of neutral null for understanding evolution. *BMC Biol.* 14:114. doi: 10.1186/s12915-016-0338-2
- Kuljanabhagavad, T., Thongphasuk, P., Chamulitrat, W., and Wink, M. (2008). Triperene saponins from *Chenopodium quinoa* Willd. *Phytochemistry* 69, 1919–1926. doi: 10.1016/j.phytochem.2008.03.001
- Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

- Lee, S. J., Park, J. H., Lee, M. H., Yu, J. H., and Kim, S. Y. (2010). Isolation and functional characterization of CE1 binding proteins. *BMC Plant Biol.* 10:277. doi: 10.1186/1471-2229-10-277
- Li, C. Y., Leopold, A. L., Sander, G. W., Shanks, J. V., Zhao, L., and Gibson, S. I. (2013). The ORCA2 transcription factor plays a key role in regulation of the terpenoid indole alkaloid pathway. *BMC Plant Biol.* 13:155. doi: 10.1186/1471-2229-13-155
- Lu, X., Zhang, L., Zhang, F., Jiang, W., Shen, Q., Zhang, L., et al. (2013). AaORC, a trichome-specific AP2/ERF transcription factor of *Artemisia annua*, is a positive regulator in the artemisinin biosynthetic pathway and in disease resistance to *Botrytis cinerea*. *New Phytol.* 198, 1191–1202. doi: 10.1111/nph.12207
- Lv, Z., Zhang, L., and Tang, K. (2017). New insights into artemisinin regulation. *Plant Signal. Behav.* 12:e1366398. doi: 10.1080/15592324.2017.1366398
- Maerkl, S. J., and Quake, S. R. (2009). Experimental determination of the evolvability of a transcription factor. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18650–18655. doi: 10.1073/pnas.0907688106
- Matsushima, R., Fukao, Y., Nishimura, M., and Hara-Nishimura, I. (2004). NAI1 gene encodes a basic helix-loop-helix-type putative transcription factor that regulates the formation of an endoplasmic reticulum-derived structure, the ER body. *Plant Cell* 16, 1536–1549. doi: 10.1105/tpc.021154
- Menke, F. L. H., Champion, A., Kijne, J. W., and Memelink, J. (1999). A novel jasmonate- and elicitor-responsive element in the periwinkle secondary metabolite biosynthetic gene *Str* interacts with a jasmonate- and elicitor-inducible AP2-domain transcription factor, ORCA2. *EMBO J.* 16, 4455–4463. doi: 10.1093/emboj/18.16.4455
- Mertens, J., Pollier, J., Vanden Bosshe, R., Lopez-Vidriero, I., Franco-Zorrilla, J. M., and Goossens, A. (2016a). The bHLH transcription factors TSAR1 and TSAR2 regulate triterpene saponin biosynthesis in *Medicago truncatula*. *Plant Physiol.* 170, 194–210. doi: 10.1104/pp.15.01645
- Mertens, J., Van Moerkercke, A., Vanden Bosshe, R., Pollier, J., and Goossens, A. (2016b). Clade IVa basic helix-loop-helix transcription factors form part of a conserved jasmonate signaling circuit for the regulation of bioactive plant terpenoid biosynthesis. *Plant Cell Physiol.* 57, 2564–2575. doi: 10.1093/pcp/pcw168
- Moghe, G. D., and Last, R. L. (2015). Something old, something new: conserved enzymes and the evolution of novelty in plant specialized metabolism. *Plant Physiol.* 169, 1512–1523. doi: 10.1104/pp.15.00994
- Nakamura, Y., Afendi, F. M., Parvin, A. K., Ono, N., Tanaka, K., Hirai, A., et al. (2014). KNApSACk metabolite activity database for retrieving the relationships between metabolites and biological activities. *Plant Cell Physiol.* 55:e7. doi: 10.1093/pcp/pct176
- Nakano, R. T., Piślewska-Bednarek, M., Yamada, K., Edger, P. P., Miyahara, M., Kondo, M., et al. (2017). PYK10 myrosinase reveals a functional coordination between endoplasmic reticulum bodies and glucosinolates in *Arabidopsis thaliana*. *Plant J.* 89, 204–210. doi: 10.1111/tpj.13377
- Nakano, R. T., Yamada, K., Bednarek, P., Nishimura, M., and Hara-Nishimura, I. (2014). ER bodies in plants of the Brassicales order: biogenesis and association with innate immunity. *Front. Plant Sci.* 5:73. doi: 10.3389/fpls.2014.00073
- Nakano, T., Suzuki, K., Fujimura, T., and Shinshi, H. (2006). Genome-wide analysis of ERF gene family in *Arabidopsis* and rice. *Plant Physiol.* 140, 411–432. doi: 10.1104/pp.105.073783
- Nakayasu, M., Shioya, N., Shikata, M., Thagun, C., Abdelkareem, A., Okabe, Y., et al. (2018). JRE4 is a master transcriptional regulator of defense-related steroidal glycoalkaloids in tomato. *Plant J.* 94, 975–990. doi: 10.1111/tpj.13911
- Nemoto, K., Takemori, N., Seki, M., Shinozaki, K., and Sawasaki, T. (2015). Members of the plant CRK superfamily are capable of trans- and autophosphorylation of tyrosine residues. *J. Biol. Chem.* 290, 16665–16677. doi: 10.1074/jbc.M114.617274
- Olofsson, L., Engstrom, A., Lundgren, A., and Brodelius, P. E. (2011). Relative expression of genes of terpene metabolism in different tissues of *Artemisia annua* L. *BMC Plant Biol.* 11:45. doi: 10.1186/1471-2229-11-45
- Osbourne, A., Goss, R. J., and Field, R. A. (2011). The saponins: polar isoprenoids with important and diverse biological activities. *Nat. Prod. Rep.* 28, 1261–1268. doi: 10.1039/c1np00015b
- Patra, B., Schluttenhofer, C., Wu, Y., Pattanaik, S., and Yuan, L. (2013). Transcriptional regulation of secondary metabolite biosynthesis in plants. *Biochim. Biophys. Acta* 1929, 1236–1247. doi: 10.1016/j.bbagr.2013.09.006
- Paul, P., Singh, S. K., Patra, B., Sui, X., Pattanaik, S., and Yuan, L. (2017). A differentially regulated AP2/ERF transcription factor gene cluster acts downstream of a MAP kinase cascade to modulate terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. *New Phytol.* 213, 1107–1123. doi: 10.1111/nph.14252
- Schweizer, F., Bodenhausen, N., Lassueur, S., Masclaux, F. G., and Reymond, F. (2013). Differential contribution of transcription factors to *Arabidopsis thaliana* defense against *Spodoptera littoralis*. *Front. Plant Sci.* 4:13. doi: 10.3389/fpls.2013.00013
- Sheard, L. B., Tan, X., Mao, H., Withers, J., Ben-Nissan, G., Hinds, T. R., et al. (2010). Jasmonate perception by inositol-phosphate-potentiated COI1-JAZ co-receptor. *Nature* 468, 400–405. doi: 10.1038/nature09430
- Shoji, T., and Hashimoto, T. (2011a). Recruitment of a duplicated primary metabolism genes into the nicotine biosynthesis regulon in tobacco. *Plant J.* 67, 949–959. doi: 10.1111/j.1365-3113.2011.04647.x
- Shoji, T., and Hashimoto, T. (2011b). Tobacco MYC2 regulates jasmonate-inducible nicotine biosynthesis genes directly and by the way of the NIC2-locus ERF genes. *Plant Cell Physiol.* 52, 1117–1130. doi: 10.1093/pcp/pcr063
- Shoji, T., and Hashimoto, T. (2015). Stress-induced expression of NICOTINE2-locus genes and their homologs encoding Ethylene Response Factor transcription factors in tobacco. *Phytochemistry* 113, 41–49. doi: 10.1016/j.phytochem.2014.05.017
- Shoji, T., and Hashimoto, T. (2019). Expression of a tobacco nicotine biosynthesis gene depends on the JRE4 transcription factor in heterogeneous tomato. *J. Plant Res.* 132, 173–180. doi: 10.1007/s10265-018-1075-0
- Shoji, T., Kajikawa, M., and Hashimoto, T. (2010). Clustered transcription factor genes regulate nicotine biosynthesis in tobacco. *Plant Cell* 22, 3390–3409. doi: 10.1105/tpc.110.078543
- Shoji, T., Mishima, M., and Hashimoto, T. (2013). Divergent DNA-binding specificities of a group of Ethylene Response Factor transcription factors involved in plant defense. *Plant Physiol.* 162, 977–990. doi: 10.1104/pp.113.217455
- Shoji, T., Nakajima, K., and Hashimoto, T. (2000). Ethylene suppresses jasmonate-induced gene expression in nicotine biosynthesis. *Plant Cell Physiol.* 41: 1072–1076. doi: 10.1093/pcp/pcd027
- Sirikantaramas, S., Asano, T., Sudo, H., Yamazaki, M., and Saito, K. (2007). Camptothecin: therapeutic potential and biotechnology. *Curr. Pharm. Biotechnol.* 8, 196–202. doi: 10.2174/138920107781387447
- Sogabe, Y., Nakamura, H., Nakagawa, T., Hasegawa, S., Asano, T., Ohta, H., et al. (2011). Plant Signal. *Behaviour* 6, 1037–1039. doi: 10.4161/psb.6.7.15602
- Srivastava, S., Chaudhry, V., Mishra, A., Chauhan, P. S., Rehman, A., Yadav, A., et al. (2012). Gene expression profiling through microarray analysis in *Arabidopsis thaliana* colonized by *Pseudomonas putida* MTC5279, a plant growth promoting rhizobacterium. *Plant Signal. Behav.* 7, 235–245. doi: 10.4161/psb.18957
- Swinnen, G., Goossens, A., and Pauwels, L. (2016). Lessons from domestication: targeting cis-regulatory elements for crop improvement. *Trends Plant Sci.* 21, 506–515. doi: 10.1016/j.tplants.2016.01.014
- Tamura, K., Yoshida, K., Hiraoka, Y., Sakaguchi, D., Chikugo, A., Mochida, K., et al. (2018). The basic helix-loop-helix transcription factor GubHLH3 positively regulates soyasaponin biosynthetic genes in *Glycyrrhiza uralensis*. *Plant Cell Physiol.* 59, 783–796. doi: 10.1093/pcp/pcy046
- Tang, K., Shen, Q., Yan, T., and Fu, X. (2014). Transgenic approach to increase artemisinin content in *Artemisia annua* L. *Plant Cell Rep.* 33, 605–615. doi: 10.1007/s00299-014-1566-y
- Thagun, C., Imanishi, S., Kudo, T., Nakabayashi, R., Ohyama, K., Mori, T., et al. (2016). Jasmonate-responsive ERF transcription factors regulate steroidal glycoalkaloid biosynthesis in tomato. *Plant Cell Physiol.* 57, 961–975. doi: 10.1093/pcp/pcw067
- Thines, B., Katsir, L., Melotto, M., Niu, Y., Mandaokar, A., Liu, G., et al. (2007). JAZ repressor proteins are targets of the SCF(COI1) complex during jasmonate signalling. *Nature* 448, 661–665. doi: 10.1038/nature05960
- Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. doi: 10.1093/nar/22.22.4673
- Udomsom, N., Rai, A., Suzuki, H., Okuyama, J., Imai, R., Mori, T., et al. (2016). Function of AP2/ERF transcription factors involved in the regulation of

- specialized metabolism in *Ophiorrhiza pumila* revealed by transcriptomics and metabolomics. *Front Plant Sci.* 7:1861. doi: 10.3389/fpls.2016.01861
- Van der Fits, L., and Memelink, J. (2000). ORCA3, a jasmonate-responsive transcriptional regulator of plant primary and secondary metabolism. *Science* 289, 295–297. doi: 10.1126/science.289.5477.295
- Van Moerkercke, A., Steensma, P., Gariboldi, I., Espoz, J., Purnama, P. C., Schweizer, F., et al. (2016). The basic helix-loop-helix transcription factor BIS2 is essential for monoterpenoid indole alkaloid production in medicinal plant *Catharanthus roseus*. *Plant J.* 88, 3–12. doi: 10.1111/tpj.13230
- Van Moerkercke, A., Steensma, P., Schweizer, F., Pollier, J., Gariboldi, I., Payne, R., et al. (2015). The bHLH transcription factor BIS1 controls the iridoid branch of the monoterpenoid indole alkaloid pathway in *Catharanthus roseus*. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8130–8135. doi: 10.1073/pnas.1504951112
- Weng, J. K., Philippe, R. N., and Noel, J. P. (2012). The rise of chemodiversity in plants. *Science* 336, 1667–1670. doi: 10.1126/science.1217411
- Wray, G. A. (2007). The evolutionary significance of cis-regulatory elements. *Nat. Rev. Genet.* 8, 206–216. doi: 10.1038/nrg2063
- Yonekura-Sakakibara, K., and Saito, K. (2013). Transcriptome coexpression analysis using ATTED-II for integrated transcriptomic/metabolomics analysis. *Methods Mol. Biol.* 1011, 317–326. doi: 10.1007/978-1-62703-414-2_25
- Zhang, F., Yao, J., Ke, J., Zhang, L., Lam, V. Q., Xin, X. F., et al. (2015). Structural basis of JAZ repression of MYC transcription factors in jasmonate signalling. *Nature* 525, 269–273. doi: 10.1038/nature14661
- Zhang, H., Hedhill, S., Montiel, G., Zhang, Y., Chatel, G., Pré, M., et al. (2011). The basic helix-loop-helix transcription factor CrMYC2 controls the jasmonate-responsive expression of the ORCA genes that regulate alkaloid biosynthesis in *Catharanthus roseus*. *Plant J.* 67, 61–71. doi: 10.1111/j.1365-313X.2011.04575.x
- Zhang, H. B., Bokowiec, M. T., Rushton, P. J., Han, S. C., and Timko, M. P. (2012). Tobacco transcription factors NtMYC2a and NtMYC2b form nuclear complexes with the NtJAZ1 repressor and regulate multiple jasmonate-inducible steps in nicotine biosynthesis. *Mol. Plant* 5, 73–84. doi: 10.1093/mp/ssr056
- Zhou, M., and Memelink, J. (2016). Jasmonate-responsive transcription factors regulating plant secondary metabolism. *Biotechnol. Adv.* 34, 441–449. doi: 10.1016/j.biotechadv.2016.02.004
- Zhu, G., Wang, S., Huang, Z., Zhang, S., Liao, Q., Zhang, C., et al. (2018). Rewiring of the fruit metabolome in tomato breeding. *Cell* 172, 249–261. doi: 10.1016/j.cell.2017.12.019
- Zhu, X., Zeng, X., Sun, C., and Chen, S. (2014). Biosynthetic pathway of terpenoid indole alkaloids in *Catharanthus roseus*. *Front Med.* 8, 285–293. doi: 10.1007/s11684-014-0350-2

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Shoji. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Chemodiversity of the Glucosinolate-Myrosinase System at the Single Cell Type Resolution

Shweta Chhajed^{1,2†}, Biswapriya B. Misra^{1,3†}, Nathalia Tello^{1,2} and Sixue Chen^{1,2,4,5*}

¹Department of Biology, University of Florida, Gainesville, FL, United States, ²Genetics Institute, University of Florida, Gainesville, FL, United States, ³Section on Molecular Medicine, Department of Internal Medicine, Center for Precision Medicine, Wake Forest School of Medicine, Winston-Salem, NC, United States, ⁴Plant Molecular and Cellular Biology, University of Florida, Gainesville, FL, United States, ⁵Interdisciplinary Center for Biotechnology Research, University of Florida, Gainesville, FL, United States

OPEN ACCESS

Edited by:

Kazuki Saito,
RIKEN Center for Sustainable
Resource Science (CSRS), Japan

Reviewed by:

Ryosuke Sugiyama,
RIKEN Center for Sustainable
Resource Science (CSRS), Japan
Franziska S. Hanschen,
Leibniz-Institut für Gemüse- und
Zierpflanzenbau (IGZ), Germany

*Correspondence:

Sixue Chen
schen@ufl.edu

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Plant Metabolism and
Chemodiversity,
a section of the journal
Frontiers in Plant Science

Received: 28 February 2019

Accepted: 25 April 2019

Published: 21 May 2019

Citation:

Chhajed S, Misra BB, Tello N and
Chen S (2019) Chemodiversity of the
Glucosinolate-Myrosinase System at
the Single Cell Type Resolution.
Front. Plant Sci. 10:618.
doi: 10.3389/fpls.2019.00618

Glucosinolates (GLSs) are a well-defined group of specialized metabolites, and like any other plant specialized metabolites, their presence does not directly affect the plant survival in terms of growth and development. However, specialized metabolites are essential to combat environmental stresses, such as pathogens and herbivores. GLSs naturally occur in many pungent plants in the order of Brassicales. To date, more than 200 different GLS structures have been characterized and their distribution differs from species to species. GLSs co-exist with classical and atypical myrosinases, which can hydrolyze GLS into an unstable aglycone thiohydroximate-O-sulfonate, which rearranges to produce different degradation products. GLSs, myrosinases, myrosinase interacting proteins, and GLS degradation products constitute the GLS-myrosinase (GM) system (“mustard oil bomb”). This review discusses the cellular and subcellular organization of the GM system, its chemodiversity, and functions in different cell types. Although there are many studies on the functions of GLSs and/or myrosinases at the tissue and whole plant levels, very few studies have focused on different single cell types. Single cell type studies will help to reveal specific functions that are missed at the tissue and organismal level. This review aims to highlight (1) recent progress in cellular and subcellular compartmentation of GLSs, myrosinases, and myrosinase interacting proteins; (2) molecular and biochemical diversity of GLSs and myrosinases; and (3) myrosinase interaction with its interacting proteins, and how it regulates the degradation of GLSs and thus the biological functions (e.g., plant defense against pathogens). Future prospects may include targeted approaches for engineering/breeding of plants and crops in the cell type-specific manner toward enhanced plant defense and nutrition.

Keywords: glucosinolate, myrosinase, cell type, metabolism, protein-protein interaction

INTRODUCTION

One of the most extensively studied classes of anti-herbivore chemical defenses in plants is glucosinolates (GLSs), a group of sulfur-rich, amino acid-derived metabolites combining a β -D-glucopyranose residue linked *via* a sulfur atom to an *N*-hydroxyimino sulfate ester, which are plant-derived natural products (Halkier and Gershenzon, 2006; Halkier, 2016).

GLSs are widely distributed in the order Brassicales, which includes vegetables (cabbage, cauliflower, and broccoli), spice plants supplying condiments (mustard, horseradish, and wasabi), and reference species, *Arabidopsis thaliana* (Fahey et al., 2001; Reichelt et al., 2002). Upon insect feeding or mechanical disruption, GLSs are hydrolyzed by myrosinases (thioglucoside glucohydrolase, TGG, EC 3.2.1.147) into unstable thiohydroximate-O-sulfonates, which rearrange to form different hydrolytic products such as isothiocyanates (ITCs), nitriles, and other by-products depending on the nature of the GLS side chain and the reaction conditions, such as iron, pH, and presence of myrosinase interacting proteins (Chen and Andreasson, 2001; Wittstock et al., 2016a). This GLS-myrosinase (GM) system is popularly known as “mustard oil bomb” (Lüthy and Matile, 1984; Ratzka et al., 2002). Myrosin cells (an idioblast cell type accumulating TGGs) are involved in plant defense by hydrolyzing GLSs into toxic volatiles such as ITCs or nitriles (Wittstock et al., 2003). TGGs are known to be present in all *A. thaliana* organs and were reported in *A. thaliana* and *B. napus* phloem parenchyma as well as in guard cells (Andréasson et al., 2001; Thangstad et al., 2004). In general, GLSs are enriched in “S-cells” that are found in *Arabidopsis* flower stalks and occur close to myrosin cells (Koroleva et al., 2000; Andréasson et al., 2001).

The spatial distribution of GLSs was demonstrated in *A. thaliana* leaves by constructing ion intensity maps from matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectra, where major GLSs were found to be more abundant in tissues of the midvein and the periphery of the leaf than the inner lamina (Shroff et al., 2008). Although this study concluded that GLSs are not abundant on *A. thaliana* leaf surfaces, the authors could not obtain information on the cell type distribution of GLSs in leaves. Moreover, all the genes in the GLS biosynthetic pathways have been identified, and it is somewhat known where GLSs are stored (Koroleva et al., 2000; Andréasson et al., 2001), but it has remained elusive where GLSs are specifically produced at the subcellular, cellular, and tissue levels (Rask et al., 2000; Nintemann et al., 2017). Neither is it clear about the cellular and subcellular compartmentation of different myrosinases and their interacting proteins, which include myrosinase-binding proteins (MBPs), myrosinase-associated proteins (MyAPs), and different specifier proteins.

In the following sections, we discuss various aspects of the GM system based on current knowledge, starting from the cellular control of enzymes, cell type, and subcellular organization, to uniqueness of myrosinases and myrosinase interacting proteins covering a range of small molecule and macromolecular interactions of the “mustard oil bomb.”

THE GLUCOSINOLATE-MYROSINASE SYSTEM AND CELLULAR CONTROL OF ENZYME REACTIONS

As found in the order of Brassicales, including important crops (e.g., mustard, oilseed rape, radish, broccoli, and cabbage), GLSs co-exist with myrosinases. When tissue damage occurs,

the “mustard oil bomb” is detonated and GLSs are hydrolyzed and converted to different degradation products with a variety of biological activities (Rask et al., 2000; Halkier and Gershenzon, 2006; Yan and Chen, 2007; Bednarek et al., 2009; Clay et al., 2009; Halkier, 2016; Wittstock et al., 2016a). For example, these degradation products play important roles in plant defense against pathogens and herbivores, as well as serve as attractants to specialists (Rask et al., 2000; Barth and Jander, 2006; Clay et al., 2009; Wittstock et al., 2016a). Several of these degradation products are involved in plant nutrition (Holmes, 1980; Armengaud et al., 2004) and growth regulation (Hasegawa et al., 2000; Hull et al., 2000; Mikkelsen et al., 2000). In plant metabolism, it is important that enzymes and substrates are under tight regulation, which is more relevant for toxic compounds, as these chemical defenses are derived from specialized metabolites. There are several ways of regulation: (1) coarse control through biosynthesis; (2) fine control of enzyme activity through protein interaction and allosteric regulation; and (3) substrate and enzyme compartmentalization (Sweetlove and Fernie, 2013). While the regulation is well studied in primary metabolism (e.g., photosynthesis and respiration), it is not clear in many of the specialized metabolic processes such as GLS metabolism. Furthermore, protein-protein interactions are intrinsic to virtually every cellular process and have been extensively studied in animals and yeast (Uetz et al., 2000; Gavin et al., 2002; Ho et al., 2002; Li et al., 2004; Huttlin et al., 2017). In plants, this area has lagged behind in spite of recent progress (Hosseinpour et al., 2012; Zhang et al., 2016; Jiang et al., 2018). Vast majority of the studies did not go beyond identifying physical interactions to the point of functional analysis. **Figure 1** shows the potential molecular interactions of the GM system in the context of cell type-specific metabolisms.

CELL TYPE-SPECIFIC CELLULAR AND SUBCELLULAR ORGANIZATION OF THE “MUSTARD OIL BOMB”

Myrosinase is located in myrosin cells, which are scattered cells in radicles, stems, leaves, petioles, seeds, and seedlings of several species (Husebye et al., 2002). A cell-specific localization was found in radicles and cotyledons of the maturing embryo resembling the pattern of the myrosin cells (Bones et al., 1991). Most GLSs are constitutively present in all *Arabidopsis* tissues (Petersen et al., 2002; Brown et al., 2003). The key steps in the biosynthesis of the different types of GLSs are localized in distinct cells in separate as well as overlapping vascular tissues (Nintemann et al., 2018). The presence of GLS biosynthetic enzymes in parenchyma cells of the vasculature may assign new defense-related functions to these cell types (Nintemann et al., 2018). To date, the cellular and subcellular compartmentation of the “mustard oil bomb” (Lüthy and Matile, 1984) is not completely clear and is rather contradictory. For instance, in *Arabidopsis* flower stalks, GLSs were found in the elongated sulfur-rich “S-cells” situated between phloem and endodermis (Koroleva et al., 2000; Husebye et al., 2002). However, the myrosinase TGG1 was found to be abundant in guard cells,

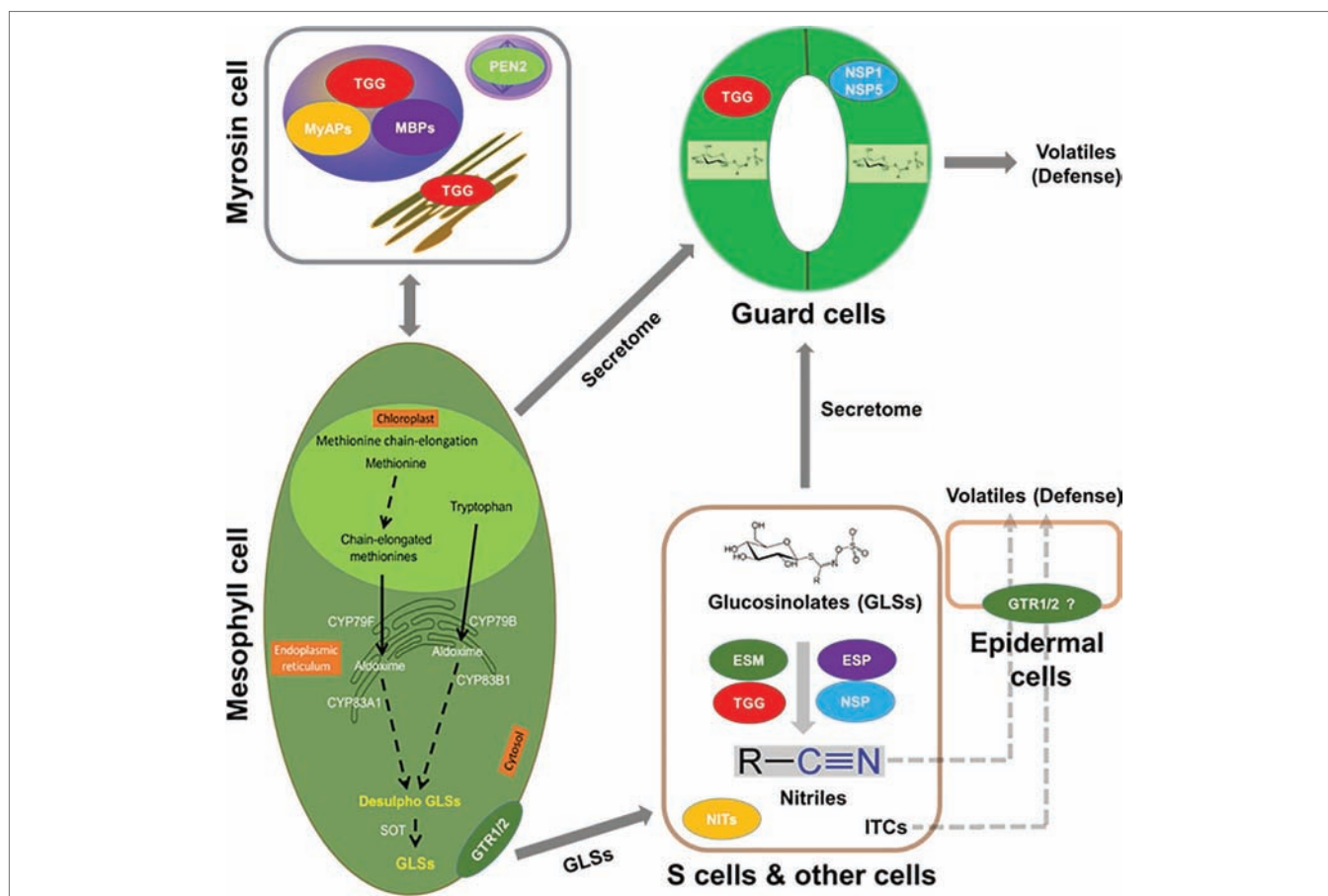


FIGURE 1 | Putative interactions between myrosinases (TGGs), myrosinase interacting proteins, GLSs, and volatiles in the context of cell type compartmentation. A myrosin cell shows vacuolar localization of TGGs, myrosinase-binding proteins (MBPs), and myrosinase-associated proteins (MyAPs); peroxisomal localization of penetration (PEN2); and ER localization of TGG. Transporters that are specific to GLSs such as NRT1/PTR glucosinolate transporter (GTR) 1, GTR2 or non-specific transporters could be aiding in their transport to site of accumulation such as S-cells or guard cells. Importantly, these cells may have the capability of *de novo* biosynthesis of GLSs. In addition, the presence of epithiospecifier modifier (ESM, MyAP-like), epithiospecifier (ESP), and nitrile specifier (NSP) 1, NSP5, etc. may lead to the breakdown of GLSs to nitriles and isothiocyanates (ITCs) for roles in cell type-specific signaling and defense against pathogen and herbivores. GLSs, TGGs, and ESP were found in the S-cells, and the presence of ESM and NSP is indicative of other cell types.

whereas TGG1 and TGG2 were localized to the phloem-associated cells close to the “S-cells” (Husebye et al., 2002; Thangstad et al., 2004; Barth and Jander, 2006). Thus, it appears that myrosinases and their substrates were physically separated in the plant tissues. However, such an arrangement may not be the case as a recent proteomics study located the myrosinases in “S-cells” (Koroleva and Cramer, 2011). In *Brassica juncea* seedlings, myrosinase was found to co-localize with GLSs in aleurone-type cells (Kelly et al., 1998). In *Arabidopsis* suspension cells, both myrosinases and GLSs were present (Alvarez et al., 2008). Such diverse co-localization results may indicate that myrosinases and GLSs are spatially separated at the subcellular levels. Alternatively, they could be in the same compartment with tight control of myrosinase activities. GLSs were found in vacuoles rich in ascorbic acid (Grob and Matile, 1979), which plays a role to inhibit myrosinase at high concentration and activate myrosinase at low concentration. This dual regulation supports the potential co-localization of GLSs and myrosinases in the same subcellular compartment.

Recent metabolomics data have confirmed the presence of GLSs in guard cells (Geng et al., 2016; Zhu and Assmann, 2017) and revealed the changes in GLS metabolism in guard cells upon treatment with CO₂ (Geng et al., 2016) and ABA (Zhu and Assmann, 2017). The first indication of the role of GLS metabolism in stomatal movement was obtained through analysis of the effect of ABA on stomatal movement of the *Arabidopsis* myrosinase mutant *tgg1* (Zhao et al., 2008). Subsequently, additional reverse genetics studies corroborated the role of GLS metabolism in stomatal movement (Islam et al., 2009; Zhu et al., 2014). Furthermore, stomatal closure was induced by pharmacological treatments with different GLS hydrolysis products (Khokon et al., 2011; Sobahan et al., 2015). However, these products and the amounts used are of synthetic origin and abundance. It is not known what degradation products are produced and how much *in vivo*, which GLSs and myrosinases [TGGs and/or Penetration 2 (PEN2)] are involved, and how protein interactions regulate the GLS breakdown in guard cells.

The *Arabidopsis cyp79b2/cyp79b3* mutants are known to produce mostly aliphatic GLSs (Zhao et al., 2002; Chen et al., 2003; Grubb and Abel, 2006; Khokon et al., 2011; Sobahan et al., 2015), while the *myb28/myb29* mutants are known to produce mostly indolic GLSs (Hirai et al., 2007; Beekwilder et al., 2008). Furthermore, the *tgg1/tgg2* double mutant showed undetectable myrosinase activity, and damage-induced breakdown of endogenous GLSs was not from aliphatic GLSs and was greatly slowed for indole GLSs (Barth and Jander, 2006). Moreover, the *tgg1/tgg2* mutant lacking the foliar myrosinases was compromised in activation of their GLS defense. Another mutant, *atvam3* mutant showed abnormal distribution of myrosin cells and overproduction of TGG1 and TGG2 (Ueda et al., 2006). Thus, beyond TGGs, MYB28, MYB29, AtVAM, CYP79s, and other biosynthetic genes all affect GLS deposition levels and possibly cell type specificity of the GM system. To understand the regulation and correlation of these proteins, we used GeneMANIA software (Warde-Farley et al., 2010) and predicted the association of the known genes involved in GLS metabolism (from our selected gene list in **Supplementary Table S1**). This software further added putative proteins with similar functions and potential involvement in the GM system (**Figure 2**).

In *B. napus* leaves, myrosinases are localized in mesophyll cells and phloem cells (Chen and Andreasson, 2001) and were mainly stored in protein-rich vacuolar structures of myrosin cells (Rask et al., 2000; Ueda et al., 2006). There is also a report of the presence of myrosinase as cytosolic enzymes bound

to intracellular membranes (Lüthy and Matile, 1984). The knowledge of the localization of myrosinases and interacting proteins was advanced by vacuolar proteomics. Myrosinases, TGG1 and TGG2, and myrosinase-associated protein (MyAP) 1 were identified in the vacuoles. In the early leaf developmental stages, TGG1 is more abundant than TGG2, whereas in fully expanded leaves, both TGG1 and TGG2 levels show increased accumulation. Concurrently, MyAP1 levels are increasingly abundant. We have previously observed such regulation of myrosinase expression, which correlated with GLS turnover (Petersen et al., 2002). The co-localization of myrosinase and MyAP1 and the concurrent expression during development lead to the hypothesis that the vacuolar myrosinases may be active and MyAPs may interact with myrosinase to play a role in GLS hydrolysis. For example, MyAPs may facilitate ITC production (Zhang et al., 2006). Indeed, immunogold analysis of leaf sections showed the presence of TGG1 and TGG2 in the same vacuoles (Ueda et al., 2006). An independent vacuolar proteomics study also identified these proteins (Carter et al., 2004). In addition, two more MyAPs (At1g54000 and At1g54010) and three myrosinase-binding proteins (MBPs) (At1g52040, At3g16470, and At2g39330) were localized in the vacuoles (Carter et al., 2004). TGG1 and TGG2 were also found in the endoplasmic reticulum (ER), ER bodies, and transvacuolar strands, and this localization is dependent on MyAP1 (MVP1). Mutation of the MyAP1 clearly altered the subcellular localization profiles of the green fluorescent protein (GFP)-tagged TGG1 and TGG2

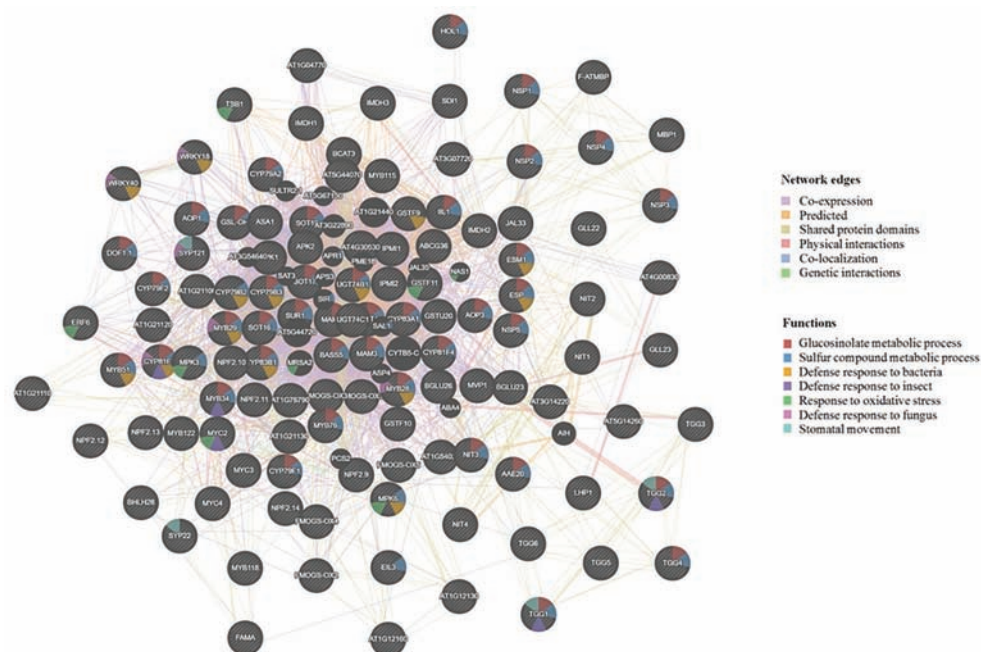


FIGURE 2 | Visualization of functional prediction of protein networks in the glucosinolate-myrosinase (GM) system using GeneMANIA (<http://genemania.org/>). The protein names are indicated inside the nodes, and the links between the nodes indicate the network edges in which the proteins are connected. The color of the edges represents evidence for the connection, which includes co-expression (purple), predicted (yellow), shared protein domains (beige), physical interactions (pink), co-localization (blue), and genetic interactions (green). As to functions associated with each protein, the color code inside the nodes indicates GLS metabolic process (red), sulfur compound metabolic process (blue), defense response to bacterium (yellow), defense response to insect (purple), response to oxidative stress (green), defense response to fungus (pink), and stomatal movement (light blue).

(Agee et al., 2010). Interestingly, the myrosinase PEN2 (hydrolyzing indole GLSs and shown to function in plant defense (Bednarek et al., 2009; Clay et al., 2009; Millet et al., 2010; Fan et al., 2011; Johansson et al., 2014; Frerigmann et al., 2016; Luti et al., 2016; Xu et al., 2016; Vilakazi et al., 2017)) is targeted to peroxisomes and the outer mitochondrial membrane (Fuchs et al., 2016). In addition to MyAPs and MBPs, specifier proteins including epithiospecifier modifier (ESM, MyAP-like), epithiospecifier protein (ESP), nitrile specifier protein (NSP), and thiocyanate forming protein (TFP) may affect the outcome of GLS degradation (Lambrix et al., 2001; Burow et al., 2006; Zhang et al., 2006; Wittstock et al., 2016a,b; Backenköhler et al., 2018). ESP was found to be in “S-cells” and in guard cells with NSP1 and NSP5 (Burow et al., 2007; Zhao et al., 2008). The functions of these MyAPs, MBPs, and specifier proteins in “S-cells” and guard cells and their interactions with myrosinases in different cell types are not known.

To understand the subcellular organization of the GM system, we compared the proteins and pathways involved in GLS biosynthesis, degradation, and transport using available and/or predicted subcellular localization information. **Figure 3** and **Supplementary Table S1** provide an overview of the GM system at subcellular level based on available literature and analysis using different protein localization tools: (1) Plant-mPLOC¹ (Chou and Shen, 2007, 2008, 2010); (2) TAIR² (with annotation based on literature); (3) Eplant³ using SUBA (Subcellular Localisation Database for *Arabidopsis*) with annotation based on subcellular proteomics and/or protein fluorescence microscopy; (4) TargetP⁴ based on the N-terminal targeting sequences (chloroplast transit peptide (cTP), mitochondrial targeting peptide (mTP), or secretory pathway signal peptide (SP) (Emanuelsson et al., 2000) [with a reliability score of 1–5 (1 being most reliable and 5 least reliable)]; (5) LocTree⁵ using support vector machines for localization prediction (in the form of expected accuracy); and (6) ngLOC⁶ using Bayesian method for prediction of localization. As shown in **Figure 3**, most GM system proteins were found to be in the cytoplasm followed by nucleus, where the transcriptional regulators were localized. All the cytochrome P450s involved in GLS biosynthesis and modification were localized to endoplasmic reticulum, and other GLS biosynthesis-related proteins were in the chloroplast and cytoplasm. Glucosinolate transporters (GTR1 and GTR2) and nitrate transporters (NRT1.6, NRT1.7, and NRT1.9) were found to be in plasma membrane. It is not known how glucosinolates are transported into vacuoles. PEN2 and BZO1 were localized in peroxisomes. No GM system proteins were found on Golgi apparatus. Out of the 114 GM system proteins used in this study (**Supplementary Table S1**), 65 proteins had experimental evidence of localization, 26 were predicted using the software tools (at least three tools with consistent result), and 23 proteins could not be conclusively localized.

DISTINCT MOLECULAR AND BIOCHEMICAL PROPERTIES OF MYROSINASES

Myrosinases are classified into two types, typical (classical) and atypical myrosinases. The crystal structure of a classical myrosinase shows that the protein folds into an (β/α)₈ barrel structure (Burmeister et al., 1997). In the active site, a Glu (E) residue is involved in nucleophilic attack to initiate the release of an aglycone (thiohydroximate-O-sulfonate) and form a glucosyl-enzyme intermediate. Another Gln (Q) residue enables the hydrolysis of this intermediate with assistance from water and ascorbate. Classical myrosinases (with QE catalytic residues) use ascorbate as a cofactor and proton donor to facilitate the release of bound glucose (Burmeister et al., 1997; Wittstock and Burow, 2010; Bhat and Vyas, 2019). In contrast, atypical myrosinases have two catalytic Glu residues (EE), which function as acid/base catalyst in the active site. They do not require ascorbate. In addition, atypical myrosinases have two basic amino acid residues at different positions (+6 and +7) for glucosinolate binding compared to +0 position arginine residue of classical myrosinases (Wittstock and Burow, 2010; Nakano et al., 2017; Shirakawa and Hara-Nishimura, 2018; Bhat and Vyas, 2019). Classical myrosinases are glycosylated, activated by low concentrations of ascorbate, and accepted GLSs as the only substrates (Chen and Halkier, 1999; Chen and Andreasson, 2001). In contrast, atypical myrosinases, such as PEN2 and PYK10, can hydrolyze indole GLSs and also use O-glucosides as substrates (Bednarek et al., 2009; Nakano et al., 2017). Although myrosinase does not use acylated GLSs and desulpho-GLSs as substrates, it may accept a wide range of GLS substrates (Chen and Halkier, 1999; Rask et al., 2000; Barth and Jander, 2006). Myrosinases from *B. napus* and *Crambe abyssinica* degrade different GLS at different rates (James and Rossiter, 1991; Finiguerra et al., 2001). However, the mechanism underlying this substrate specificity is not established. Myrosinases in *B. napus* are encoded by >29 genes in three subfamilies, denoted as MA, MB, and MC. The MA myrosinases occur as dimers, while MB and MC myrosinases exist in complexes with MBPs and/or MyAPs (Lenman et al., 1990; Rask et al., 2000). By heterologous expression in yeast, we have previously produced a functional free form myrosinase Myr1 from the MB subfamily (Chen and Halkier, 1999). The activity of this Myr1 suggests that MBPs and MyAPs are not absolutely necessary for myrosinase activity, but raises questions on the functions of MBPs and MyAPs and their interactions with myrosinases.

Bioinformatic analysis of the *Arabidopsis* genome revealed the presence of six myrosinase genes *TGG1-TGG6* (Xu et al., 2004). *TGG1* and *TGG2* are expressed in leaves (Xue et al., 1995; Husebye et al., 2002; Thangstad et al., 2004; Barth and Jander, 2006; Ueda et al., 2006) and flowers (Ruan et al., 1998; Barth and Jander, 2006), while *TGG4* and *TGG5* are specifically expressed in roots (Zimmermann et al., 2004). *TGG3* and *TGG6* are pseudogenes (Husebye et al., 2002; Zhang et al., 2002). Although *TGG1* and *TGG2* appear to display a low degree of substrate specificity, the activities of *TGG1* and

¹<http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/>

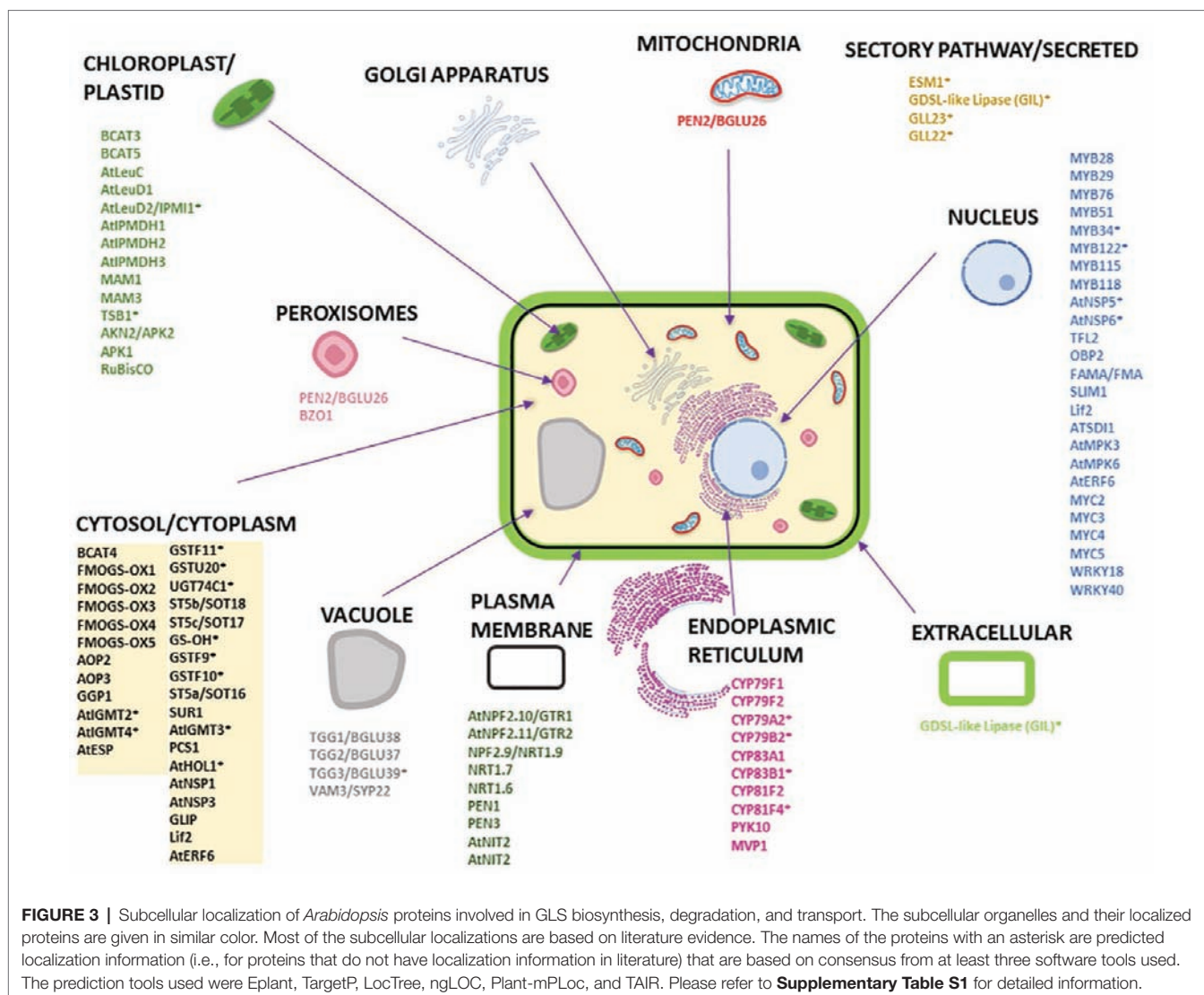
²<https://www.arabidopsis.org/tools/bulk/protein/index.jsp>

³<https://bar.utoronto.ca/eplant/>

⁴<http://www.cbs.dtu.dk/services/TargetP/>

⁵<https://roslab.org/services/loctree3/>

⁶<http://genome.unmc.edu/ngLOC/index.html>



TGG2 have been correlated with the feeding preference and growth of different generalist and specialist insects (Barth and Jander, 2006). Interestingly, overexpression of TGG1 and TGG2 leads to accumulation of several GLS degradation products, including 5-methylhexanenitrile, heptanenitrile, 1-isothiocyanato-3-methylbutane, 1-isothiocyanato-4-methylpentane, and 1-isothiocyanato-3-methylhexane. Based on the degradation product profile, possible endogenous substrates for the two TGGs include 4-methylthiobutylglucosinolate, 4-methylpentylglucosinolate and 3-methylbutylglucosinolates (Ueda et al., 2006). Investigating endogenous substrates of different classical and atypical myrosinases is an important future direction.

In leaves, TGG1 was found to be abundant in guard cells, while TGG2 appeared only present in phloem-associated cells (Barth and Jander, 2006; Zhao et al., 2008). Considering the presence of GLSs in guard cells (Geng et al., 2016; Zhu and Assmann, 2017), how the GM system plays a role

in guard cell functions (e.g., stomatal immunity) is an interesting question. Clearly, mutation of the TGG1 and/or TGG2 genes affected the guard cell size, stomatal aperture, and leaf metabolites, such as fatty acids, glucosinolates, and indole compounds (Ahuja et al., 2016). Another proteomic study of trichome and epidermal pavement cells did not identify the TGG1 protein in the samples (Wienkoop et al., 2004). However, a single cell type study in trichomes found the presence of gene encoding transcription factors of aliphatic GLS (MYB28, MYB29 and MYB76) and indole GLS (MYB34, MYB51 and MYB122), indicating that trichomes have biosynthetic genes for the GM system (Frerigmann et al., 2012), but nothing was suggested about myrosinases activity or expression. Given the defense roles of guard cells and trichomes, characterization of the GM systems in these special cell types is of great importance to understand the molecular mechanisms underlying the cell type-specific functions, e.g., defense against pathogen invasion.

COMPLEX FORMATION BETWEEN MYROSINASE AND ITS INTERACTING PROTEINS

As described earlier, myrosinase interacting proteins include MBP, MyAP, and specifier proteins (ESM, ESP, NSP, and TFP). The first six MBPs identified in *B. napus* range in size from 30 to 110 kDa (Taipalensuu et al., 1996; Chisholm et al., 2000). All MBPs contain jacalin-like repeats (Chisholm et al., 2000; Andréasson et al., 2001). Jacalin-related proteins share the domain structure of plant lectins and are upregulated by phytohormones (e.g., jasmonic acid, salicylic acid, and ethylene) and pathogens (Taipalensuu et al., 1996; Geshi and Brandt, 1998; Xiang et al., 2011; Vilakazi et al., 2017). Recently, identification of bacterial lipopolysaccharide interacting proteins in *Arabidopsis* revealed myrosinases, TGG1 and TGG2, and a MBP (Vilakazi et al., 2017). It remains unclear how the MBP levels are regulated and whether MBPs directly interact and affect myrosinase activity and specificity. In *B. napus* seeds, MBPs are present in most cells but not in the myrosin cells (Rask et al., 2000; Ueda et al., 2006). During germination, MBPs are co-localized with myrosinases in cotyledons, suggesting that preformed myrosinase complexes do exist (Geshi and Brandt, 1998; Eriksson et al., 2002). Using basic local alignment search tool (BLAST) to interrogate the *Arabidopsis* genome reveals >30 putative MBPs. MBP1 and MBP2 are like lectin jacalins and plant aggregating factors. MBP1 and MBP2 are abundantly expressed in immature flowers, and the pattern is similar to that of myrosinase TGG1 (Capella et al., 2001). MBP expression and myrosinase activity are affected in the *coil* mutant, which is insensitive to jasmonate (Capella et al., 2001). Depletion of MBPs does not alter the cellular distribution of myrosinases but prevents myrosinases from forming complexes (Eriksson et al., 2002). Thus, the functions of MBPs are not fully understood. Interestingly, most NSPs possess jacalin-like domains and are MBP-like (Kuchernig et al., 2012). The jacalin-like domain may interact with the glycans of myrosinases to potentially affect GLS degradation. However, experimental evidence is lacking. NSPs were shown to enhance simple nitrile formation (He et al., 2009; Kissen and Bones, 2009; Chen et al., 2015; Wittstock et al., 2016b). Recently, iron was shown to be a centrally bound cofactor of ESP, TFP, and NSP involved in glucosinolate breakdown. In addition, NSP active site has fewer restrictions to the aglycone conformation than ESP and TFP. This may explain why NSP facilitates simple nitrile production, but not production of epithionitrile and thiocyanate that may need exact positioning of the aglycone thiolate relative to the side chain (Backenköhler et al., 2018). In addition to MBPs, MyAPs form complexes with myrosinases in *B. napus* (Taipalensuu et al., 1996). In *Arabidopsis*, TGG2 was pulled down with MyAP1 in leaf extracts (Agee et al., 2010). MyAPs display high similarity to GDSL lipases, which have a motif of Gly, Asp, Ser, and Leu residues in the active site. The *Arabidopsis* genome contains >80 genes encoding GDSL lipases, typically with a GDSL-like motif, a catalytic triad of Ser, Asp and His residues, and a lipase signature sequence GxSxxxG

(Brick et al., 1995). The possible lipase activity of MyAP suggests a potential role of MyAP in releasing acyl groups from acylated GLSs, thereby making them available for myrosinase hydrolysis. *Arabidopsis* contains acylated GLSs in seeds, but *B. napus* does not contain acylated GLSs; thus, MyAP in *B. napus* may have other functions. A recent study shows that overexpression of *B. napus* MyAP1 led to enhanced plant defense against a fungal pathogen *Sclerotinia sclerotiorum* (Wu et al., 2017). A MyAP-like ESM was found to favor ITC production and protect *Arabidopsis* from herbivory (Zhang et al., 2006). However, whether this system involves myrosinase complex formation is still not known. In some plants, ESPs are involved in GLS hydrolysis (Foo et al., 2000; Burow et al., 2006). Hydrolysis of alkenyl GLSs in the presence of ESP leads to the formation of nitriles or epithionitriles, instead of isothiocyanates (Zabala Mde et al., 2005; Burow et al., 2006). Because ESPs can alter the course of hydrolysis, they are important in determining plant herbivore choice and host resistance (Lambrix et al., 2001). Furthermore, this suggests that ESP is situated close to the active site so that it could promptly convert the unstable aglycone to nitriles. Although kinetic studies have showed that ESP acts as a non-competitive inhibitor of myrosinase (MacLeod and Rossiter, 1985), no stable interaction between ESP and myrosinase has been reported (Burow et al., 2006). Like nitrile formation, the production of thiocyanate was found to be associated with TFP. For detailed description of myrosinase specifier proteins, please refer to a recent review (Wittstock et al., 2016a). In summary, several other groups of proteins may interact with myrosinases and function to affect how GLSs are degraded, leading to the formation of different metabolic products. Systematic studies to characterize the interaction of these proteins with myrosinases are needed to elucidate their specific functions.

DIRECTIONS FOR FUTURE RESEARCH AND CONCLUSIONS

Currently, the cellular and subcellular location of myrosinases, GLSs, and their interacting proteins, i.e., the GM system, are far from established. Given >100 cell types in plants and >5,500 species of GLS producers, it would be a challenge to capture all the species-specific and cell type-specific information of the “mustard oil bomb.” In addition, with temporal accumulation and expression patterns of metabolites and enzymes involved typically in case of specialized metabolites, these eventual pictures could be very complex. Using cell type-specific genetic manipulations (e.g., GFP fusion and CRISPR), one can envision to capture the cell type-specific expression patterns and functional role of the glucosinolate biosynthetic proteins, myrosinases, and myrosinase interacting proteins. There exist large gaps in the knowledge base of the GM system, e.g., myrosinase interacting proteins in terms of their interactions, co-localizations, regulations, and functions in specific cell types. Furthermore, the developmental staged appearance and regulation

of the proteins and metabolites are not clear. Moreover, the reported interactions of myrosinases and other proteins could be very much cell type-specific or subcellular localized, which is not well studied till date. Without resolving the cell type specificity of the proteins and metabolites, it would be very challenging to draw mechanistic conclusions on the specific roles of the enzymes, interactors, transporters, and the metabolites from tissue- and whole plant-based data where the information are averaged out (Dai and Chen, 2012; Misra et al., 2014).

In the future, efforts need to focus on large-scale speedy preparations of organelles and subcellular fractions (e.g., vacuoles, peroxisomes, and chloroplasts) in a time-dependent manner to capture the dynamics of protein interactions and GLS metabolism. It is obviously challenging to prepare and enrich plant cell types (e.g., the “S-cells”) in copious amounts for more system-wide experiments such as transcriptomics, proteomics, and metabolomics and to obtain preparations at a given time and for a specific treatment. With the recent development of single-cell omics tools (Misra et al., 2014; Efroni and Birnbaum, 2016; Doerr, 2019), such large-scale molecular characterization of different single cells is within sight and will greatly enhance the understanding of the chemodiversity of the GM system at the single-cell resolution.

REFERENCES

- Agee, A. E., Surpin, M., Sohn, E. J., Girke, T., Rosado, A., Kram, B. W., et al. (2010). MODIFIED VACUOLE PHENOTYPE1 is an *Arabidopsis* myrosinase-associated protein involved in endomembrane protein trafficking. *Plant Physiol.* 152, 120–132. doi: 10.1104/pp.109.145078
- Ahuja, I., de Vos, R. C., Rohloff, J., Stoop, G. M., Halle, K. K., Ahmad, S. J., et al. (2016). *Arabidopsis* myrosinases link the glucosinolate-myrosinase system and the cuticle. *Sci. Rep.* 6:38990. doi: 10.1038/srep38990
- Alvarez, S., He, Y., and Chen, S. (2008). Comparative investigations of the glucosinolate-myrosinase system in *Arabidopsis* suspension cells and hypocotyls. *Plant Cell Physiol.* 49, 324–333. doi: 10.1093/pcp/pcn007
- Andréasson, E., Jørgensen, L. B., Höglund, A. S., Rask, L., and Meijer, J. (2001). Different myrosinase and idioblast distribution in *Arabidopsis* and *Brassica napus*. *Plant Physiol.* 127, 1750–1763. doi: 10.1104/pp.010334
- Armengaud, P., Breitling, R., and Amtmann, A. (2004). The potassium-dependent transcriptome of *Arabidopsis* reveals a prominent role of jasmonic acid in nutrient signaling. *Plant Physiol.* 136, 2556–2576. doi: 10.1104/pp.104.046482
- Backenköhler, A., Eisenschmidt, D., Schneegans, N., Strieker, M., Brandt, W., and Wittstock, U. (2018). Iron is a centrally bound cofactor of specifier proteins involved in glucosinolate breakdown. *PLoS One* 13:e0205755. doi: 10.1371/journal.pone.0205755
- Barth, C., and Jander, G. (2006). *Arabidopsis* myrosinases TGG1 and TGG2 have redundant function in glucosinolate breakdown and insect defense. *Plant J.* 46, 549–562. doi: 10.1111/j.1365-3113.2006.02716.x
- Bednarek, P., Piślewska-Bednarek, M., Svatoš, A., Schneider, B., Doubek, J., Mansurova, M., et al. (2009). A glucosinolate metabolism pathway in living plant cells mediates broad-spectrum antifungal defense. *Science* 323, 101–106. doi: 10.1126/science.1163732
- Beekwilder, J., Van Leeuwen, W., Van Dam, N. M., Bertossi, M., Grandi, V., Mizzi, L., et al. (2008). The impact of the absence of aliphatic glucosinolates on insect herbivory in *Arabidopsis*. *PLoS One* 3:e2068. doi: 10.1371/journal.pone.0002068
- Bhat, R., and Vyas, D. (2019). Myrosinase: insights on structural, catalytic, regulatory, and environmental interactions. *Crit. Rev. Biotechnol.* 39, 508–523. doi: 10.1080/07388551.2019.1576024

AUTHOR CONTRIBUTIONS

SChh made the list of protein localization, **Figures 2, 3**, generated the reference list, and edited the manuscript. BM wrote one-third of the manuscript draft, made **Figure 1** draft, and edited the manuscript. NT assisted with **Supplementary Table S1** analysis, reference list, and edited the manuscript. SChe designed the manuscript, wrote two-third of the text, provided guidance to students, and finalized the manuscript for submission.

FUNDING

The authors would like to acknowledge funding from the National Science Foundation MCB 1758820 to SChe. A NIH supported SF2UF Bridge Program (to Dr. David Julian) and a NSF REU grant 1560049 (to Dr. Ramesh Katam) have provided support to NT.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.00618/full#supplementary-material>

- Bones, A. M., Thangstad, O. P., Haugen, O. A., and Espevik, T. (1991). Fate of myrosin cells: characterization of monoclonal antibodies against myrosinase. *J. Exp. Bot.* 42, 1541–1550. doi: 10.1093/jxb/42.12.1541
- Brick, D. J., Brumlik, M. J., Buckley, J. T., Cao, J. X., Davies, P. C., Misra, S., et al. (1995). A new family of lipolytic plant enzymes with members in rice, *Arabidopsis* and maize. *FEBS Lett.* 377, 475–480. doi: 10.1016/0014-5793(95)01405-5
- Brown, P. D., Tokuhisa, J. G., Reichelt, M., and Gershenzon, J. (2003). Variation of glucosinolate accumulation among different organs and developmental stages of *Arabidopsis thaliana*. *Phytochemistry* 62, 471–481. doi: 10.1016/S0031-9422(02)00549-6
- Burmeister, W. P., Cottaz, S., Driguez, H., Iori, R., Palmieri, S., and Henrissat, B. (1997). The crystal structures of *Sinapis alba* myrosinase and a covalent glycosyl-enzyme intermediate provide insights into the substrate recognition and active-site machinery of an S-glycosidase. *Structure* 5, 663–676. doi: 10.1016/S0969-2126(97)00221-9
- Burow, M., Markert, J., Gershenzon, J., and Wittstock, U. (2006). Comparative biochemical characterization of nitrile-forming proteins from plants and insects that alter myrosinase-catalysed hydrolysis of glucosinolates. *FEBS J.* 273, 2432–2446. doi: 10.1111/j.1742-4658.2006.05252.x
- Burow, M., Rice, M., Hause, B., Gershenzon, J., and Wittstock, U. (2007). Cell- and tissue-specific localization and regulation of the epithiospecifier protein in *Arabidopsis thaliana*. *Plant Mol. Biol.* 64, 173–185. doi: 10.1007/s11103-007-9143-1
- Capella, A., Menossi, M., Arruda, P., and Benedetti, C. (2001). COI1 affects myrosinase activity and controls the expression of two flower-specific myrosinase-binding protein homologues in *Arabidopsis*. *Planta* 213, 691–699. doi: 10.1007/s004250100548
- Carter, C., Pan, S., Zouhar, J., Avila, E. L., Girke, T., and Raikhel, N. V. (2004). The vegetative vacuole proteome of *Arabidopsis thaliana* reveals predicted and unexpected proteins. *Plant Cell* 16, 3285–3303. doi: 10.1105/tpc.104.027078
- Chen, S., and Andréasson, E. (2001). Update on glucosinolate metabolism and transport. *Plant Physiol. Biochem.* 39, 743–758. doi: 10.1016/S0981-9428(01)01301-8
- Chen, Y., Fei, M., Wang, Y., Chen, S., and Yan, X. (2015). Proteomic investigation of glucosinolate systematically changes in *Arabidopsis* rosette leaves to exogenous methyl jasmonate. *Plant Biosyst.* 149, 346–353. doi: 10.1080/11263504.2013.819044

- Chen, S., Glawischnig, E., Jørgensen, K., Naur, P., Jørgensen, B., Olsen, C. E., et al. (2003). CYP79F1 and CYP79F2 have distinct functions in the biosynthesis of aliphatic glucosinolates in *Arabidopsis*. *Plant J.* 33, 923–937. doi: 10.1046/j.1365-313X.2003.01679.x
- Chen, S., and Halkier, B. A. (1999). Functional expression and characterization of the myrosinase MYR1 from *Brassica napus* in *Saccharomyces cerevisiae*. *Protein Expr. Purif.* 17, 414–420. doi: 10.1006/prep.1999.1158
- Chisholm, S. T., Mahajan, S. K., Whitham, S. A., Yamamoto, M. L., and Carrington, J. C. (2000). Cloning of the *Arabidopsis* RTM1 gene, which controls restriction of long-distance movement of tobacco etch virus. *Proc. Natl. Acad. Sci. USA* 97, 489–494. doi: 10.1073/pnas.97.1.489
- Chou, K. C., and Shen, H. B. (2007). Large-scale plant protein subcellular location prediction. *J. Cell. Biochem.* 100, 665–678. doi: 10.1002/jcb.21096
- Chou, K. C., and Shen, H. B. (2008). Cell-PLoc: a package of web servers for predicting subcellular localization of proteins in various organisms. *Nat. Protoc.* 3:153. doi: 10.1038/nprot.2007.494
- Chou, K. C., and Shen, H. B. (2010). Plant-mPLoc: a top-down strategy to augment the power for predicting plant protein subcellular localization. *PLoS One* 5:e11335. doi: 10.1371/journal.pone.0011335
- Clay, N. K., Adio, A. M., Denoux, C., Jander, G., and Ausubel, F. M. (2009). Glucosinolate metabolites required for an *Arabidopsis* innate immune response. *Science* 323, 95–101. doi: 10.1126/science.1164627
- Dai, S., and Chen, S. (2012). Single-cell-type proteomics toward a holistic understanding of plant functions. *Mol. Cell. Proteomics* 11, 1622–1630. doi: 10.1074/mcp.R112.021550
- Doerr, A. (2019). Single-cell proteomics. *Nat. Methods* 16:20. doi: 10.1038/s41592-018-0273-y
- Efroni, I., and Birnbaum, K. D. (2016). The potential of single-cell profiling in plants. *Genome Biol.* 17:65. doi: 10.1186/s13059-016-0931-2
- Emanuelsson, O., Nielsen, H., Brunak, S., and Von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* 300, 1005–1016. doi: 10.1006/jmbi.2000.3903
- Eriksson, S., Andréasson, E., Ekbom, B., Granér, G., Pontoppidan, B., Taipaleenmäki, J., et al. (2002). Complex formation of myrosinase isoenzymes in oilseed rape seeds are dependent on the presence of myrosinase-binding proteins. *Plant Physiol.* 129, 1592–1599. doi: 10.1104/pp.003285
- Fahey, J. W., Zalcmann, A. T., and Talalay, P. (2001). The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. *Phytochemistry* 56, 5–51. doi: 10.1016/S0031-9422(00)00316-2
- Fan, J., Crooks, C., Creissen, G., Hill, L., Fairhurst, S., Doerner, P., et al. (2011). *Pseudomonas* sax genes overcome aliphatic isothiocyanate-mediated non-host resistance in *Arabidopsis*. *Science* 331, 1185–1188. doi: 10.1126/science.1199707
- Finiguerra, M. G., Iori, R., and Palmieri, S. (2001). Soluble and total myrosinase activity in defatted Crambe abyssinica meal. *J. Agric. Food Chem.* 49, 840–845. doi: 10.1021/jf000917h
- Foo, H. L., Grønning, L. M., Goodenough, L., Bones, A. M., Danielsen, B. E., Whiting, D. A., et al. (2000). Purification and characterisation of epithiospecific protein from *Brassica napus*: enzymic intramolecular sulphur addition within alkenyl thiohydroximates derived from alkenyl glucosinolate hydrolysis. *FEBS Lett.* 468, 243–246. doi: 10.1016/S0014-5793(00)01176-5
- Frerigmann, H., Böttcher, C., Baatout, D., and Gigolashvili, T. (2012). Glucosinolates are produced in trichomes of *Arabidopsis thaliana*. *Front. Plant Sci.* 3:242. doi: 10.3389/fpls.2012.00242
- Frerigmann, H., Piślewska-Bednarek, M., Sánchez-Vallet, A., Molina, A., Glawischnig, E., Gigolashvili, T., et al. (2016). Regulation of pathogen-triggered tryptophan metabolism in *Arabidopsis thaliana* by MYB transcription factors and indole glucosinolate conversion products. *Mol. Plant* 9, 682–695. doi: 10.1016/j.molp.2016.01.006
- Fuchs, R., Kopischke, M., Klapprodt, C., Hause, G., Meyer, A. J., Schwarzländer, M., et al. (2016). Immobilized subpopulations of leaf epidermal mitochondria mediate PENETRATION2-dependent pathogen entry control in *Arabidopsis*. *Plant Cell* 28, 130–145. doi: 10.1105/tpc.15.00887
- Gavin, A. C., Bösch, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., et al. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415, 141–147. doi: 10.1038/415141a
- Geng, S., Misra, B. B., De Armas, E., Huhman, D. V., Alborn, H. T., Sumner, L. W., et al. (2016). Jasmonate-mediated stomatal closure under elevated CO₂ revealed by time-resolved metabolomics. *Plant J.* 88, 947–962. doi: 10.1111/tj.13296
- Geshi, N., and Brandt, A. (1998). Two jasmonate-inducible myrosinase-binding proteins from *Brassica napus* L. seedlings with homology to jacalin. *Planta* 204, 295–304. doi: 10.1007/s004250050259
- Grob, K., and Matile, P. H. (1979). Vacuolar location of glucosinolates in horseradish root cells. *Plant Sci. Lett.* 14, 327–335. doi: 10.1016/S0304-4211(79)90281-5
- Grubb, C. D., and Abel, S. (2006). Glucosinolate metabolism and its control. *Trends Plant Sci.* 11, 89–100. doi: 10.1016/j.tplants.2005.12.006
- Halkier, B. A. (2016). “General introduction to glucosinolates” in *Advances in Botanical Research*. Vol. 80, ed. S. Kopriva (Cambridge, MA, USA: Academic Press), 1–14. doi: 10.1016/bs.abr.2016.07.001
- Halkier, B. A., and Gershenzon, J. (2006). Biology and biochemistry of glucosinolates. *Annu. Rev. Plant Biol.* 57, 303–333. doi: 10.1146/annurev.arplant.57.032905.105228
- Hasegawa, T., Yamada, K., Kosemura, S., Yamamura, S., and Hasegawa, K. (2000). Phototropic stimulation induces the conversion of glucosinolate to phototropism-regulating substances of radish hypocotyls. *Phytochemistry* 54, 275–279. doi: 10.1016/S0031-9422(00)00080-7
- He, Y., Mawhinney, T. P., Preuss, M. L., Schroeder, A. C., Chen, B., Abraham, L., et al. (2009). A redox-active isopropylmalate dehydrogenase functions in the biosynthesis of glucosinolates and leucine in *Arabidopsis*. *Plant J.* 60, 679–690. doi: 10.1111/j.1365-313X.2009.03990.x
- Hirai, M. Y., Sugiyama, K., Sawada, Y., Tohge, T., Obayashi, T., Suzuki, A., et al. (2007). Omics-based identification of *Arabidopsis* Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *Proc. Natl. Acad. Sci. USA* 104, 6478–6483. doi: 10.1073/pnas.0611629104
- Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D., Moore, L., Adams, S. L., et al. (2002). Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* 415, 180–183. doi: 10.1038/415180a
- Holmes, M. R. J. (1980). *Nutrition of the oilseed rape crop*. (London, UK: Applied Science Publishers).
- Hosseinpour, B., HajiHoseini, V., Kashfi, R., Ebrahimie, E., and Hemmatzadeh, F. (2012). Protein interaction network of *Arabidopsis thaliana* female gametophyte development identifies novel proteins and relations. *PLoS One* 7:e49931. doi: 10.1371/journal.pone.0049931
- Hull, A. K., Vij, R., and Celenza, J. L. (2000). *Arabidopsis* cytochrome P450s that catalyze the first step of tryptophan-dependent indole-3-acetic acid biosynthesis. *Proc. Natl. Acad. Sci. USA* 97, 2379–2384. doi: 10.1073/pnas.040569997
- Husebye, H., Chadchawan, S., Winge, P., Thangstad, O. P., and Bones, A. M. (2002). Guard cell- and phloem idioblast-specific expression of thioglucoside glucosyltransferase 1 (myrosinase) in *Arabidopsis*. *Plant Physiol.* 128, 1180–1188. doi: 10.1104/pp.010925
- Huttl, E. L., Bruckner, R. J., Paulo, J. A., Cannon, J. R., Ting, L., Baltier, K., et al. (2017). Architecture of the human interactome defines protein communities and disease networks. *Nature* 545, 505–509. doi: 10.1038/nature22366
- Islam, M. M., Tani, C., Watanabe-Sugimoto, M., Uraji, M., Jahan, M. S., Masuda, C., et al. (2009). Myrosinases, TGG1 and TGG2, redundantly function in ABA and MeJA signaling in *Arabidopsis* guard cells. *Plant Cell Physiol.* 50, 1171–1175. doi: 10.1093/pcp/pcp066
- James, D. C., and Rossiter, J. T. (1991). Development and characteristics of myrosinase in *Brassica napus* during early seedling growth. *Physiol. Plant.* 82, 163–170. doi: 10.1111/j.1399-3054.1991.tb00076.x
- Jiang, M., Niu, C., Cao, J., Ni, D. A., and Chu, Z. (2018). In silico-prediction of protein–protein interactions network about MAPKs and PP2Cs reveals a novel docking site variants in *Brachypodium distachyon*. *Sci. Rep.* 8:15083. doi: 10.1038/s41598-018-33428-5
- Johansson, O. N., Fantozzi, E., Fahlberg, P., Nilsson, A. K., Buhot, N., Tör, M., et al. (2014). Role of the penetration-resistance genes PEN 1, PEN 2 and PEN 3 in the hypersensitive response and race-specific resistance in *Arabidopsis thaliana*. *Plant J.* 79, 466–476. doi: 10.1111/tj.12571
- Kelly, P. J., Bones, A., and Rossiter, J. T. (1998). Sub-cellular immunolocalization of the glucosinolate sinigrin in seedlings of *Brassica juncea*. *Planta* 206, 370–377. doi: 10.1007/s004250050412
- Khokan, M. A. R., Jahan, M. S., Rahman, T., Hossain, M. A., Muroyama, D., Minami, I., et al. (2011). Allyl isothiocyanate (AITC) induces stomatal closure in *Arabidopsis*. *Plant Cell Environ.* 34, 1900–1906. doi: 10.1111/j.1365-3040.2011.02385.x

- Kissen, R., and Bones, A. M. (2009). Nitrile-specifier proteins involved in glucosinolate hydrolysis in *Arabidopsis thaliana*. *J. Biol. Chem.* 284, 12057–12070. doi: 10.1074/jbc.M807500200
- Koroleva, O. A., and Cramer, R. (2011). Single-cell proteomic analysis of glucosinolate-rich S-cells in *Arabidopsis thaliana*. *Methods* 54, 413–423. doi: 10.1016/j.ymeth.2011.06.005
- Koroleva, O. A., Davies, A., Deeken, R., Thorpe, M. R., Tomos, A. D., and Hedrich, R. (2000). Identification of a new glucosinolate-rich cell type in *Arabidopsis* flower stalk. *Plant Physiol.* 124, 599–608. doi: 10.1104/pp.124.2.599
- Kuchernig, J. C., Burow, M., and Wittstock, U. (2012). Evolution of specifier proteins in glucosinolate-containing plants. *BMC Evol. Biol.* 12:127. doi: 10.1186/1471-2148-12-127
- Lambrix, V., Reichelt, M., Mitchell-Olds, T., Kliebenstein, D. J., and Gershenzon, J. (2001). The *Arabidopsis* epithiospecifier protein promotes the hydrolysis of glucosinolates to nitriles and influences *Trichoplusia ni* herbivory. *Plant Cell* 13, 2793–2807. doi: 10.1105/tpc.13.12.2793
- Lenman, M., Rödin, J., Josefsson, L. G., and Rask, L. (1990). Immunological characterization of rapeseed myrosinase. *Eur. J. Biochem.* 194, 747–753. doi: 10.1111/j.1432-1033.1990.tb19465.x
- Li, S., Armstrong, C. M., Bertin, N., Ge, H., Milstein, S., Boxem, M., et al. (2004). A map of the interactome network of the metazoan *C. elegans*. *Science* 303, 540–543. doi: 10.1126/science.1091403
- Lüthy, B., and Matile, P. (1984). The mustard oil bomb: rectified analysis of the subcellular organisation of the myrosinase system. *Biochem. Physiol. Pflanz.* 179, 5–12. doi: 10.1016/S0015-3796(84)80059-1
- Luti, S., Caselli, A., Taiti, C., Bazihizina, N., Gonnelli, C., Mancuso, S., et al. (2016). PAMP activity of cerato-platanin during plant interaction: an-omic approach. *Int. J. Mol. Sci.* 17:866. doi: 10.3390/ijms17060866
- MacLeod, A. J., and Rossiter, J. T. (1985). The occurrence and activity of epithiospecifier protein in some cruciferae seeds. *Phytochemistry* 24, 1895–1898. doi: 10.1016/S0031-9422(00)83087-3
- Mikkelsen, M. D., Hansen, C. H., Wittstock, U., and Halkier, B. A. (2000). Cytochrome P450 CYP79B2 from *Arabidopsis* catalyzes the conversion of tryptophan to indole-3-acetaldoxime, a precursor of indole glucosinolates and indole-3-acetic acid. *J. Biol. Chem.* 275, 33712–33717. doi: 10.1074/jbc.M001667200
- Millet, Y. A., Danna, C. H., Clay, N. K., Songnuan, W., Simon, M. D., Werck-Reichhart, D., et al. (2010). Innate immune responses activated in *Arabidopsis* roots by microbe-associated molecular patterns. *Plant Cell* 22, 973–990. doi: 10.1105/tpc.109.069658
- Misra, B. B., Assmann, S. M., and Chen, S. (2014). Plant single cell and single cell-type metabolomics. *Trends Plant Sci.* 19, 637–646. doi: 10.1016/j.tplants.2014.05.005
- Nakano, R. T., Piślewska-Bednarek, M., Yamada, K., Edger, P. P., Miyahara, M., Kondo, M., et al. (2017). PYK10 myrosinase reveals a functional coordination between endoplasmic reticulum bodies and glucosinolates in *Arabidopsis thaliana*. *Plant J.* 89, 204–220. doi: 10.1111/tpl.13377
- Nintemann, S. J., Hunziker, P., Andersen, T. G., Schulz, A., Burow, M., and Halkier, B. A. (2018). Localization of the glucosinolate biosynthetic enzymes reveals distinct spatial patterns for the biosynthesis of indole and aliphatic glucosinolates. *Physiol. Plant.* 163, 138–154. doi: 10.1111/ppl.12672
- Nintemann, S. J., Vik, D., Svozil, J., Bak, M., Baerenfaller, K., Burow, M., et al. (2017). Unravelling protein-protein interaction networks linked to aliphatic and indole glucosinolate biosynthetic pathways in *Arabidopsis*. *Front. Plant Sci.* 8:2028. doi: 10.3389/fpls.2017.02028
- Petersen, B., Chen, S., Hansen, C., Olsen, C., and Halkier, B. (2002). Composition and content of glucosinolates in developing *Arabidopsis thaliana*. *Planta* 214, 562–571. doi: 10.1007/s004250100659
- Rask, L., Andréasson, E., Ekblom, B., Eriksson, S., Pontoppidan, B., and Meijer, J. (2000). Myrosinase: gene family evolution and herbivore defense in Brassicaceae. *Plant Mol. Biol.* 42, 93–114. doi: 10.1023/A:1006380021658
- Ratzka, A., Vogel, H., Kliebenstein, D. J., Mitchell-Olds, T., and Kroymann, J. (2002). Disarming the mustard oil bomb. *Proc. Natl. Acad. Sci. USA* 99, 11223–11228. doi: 10.1073/pnas.172112899
- Reichelt, M., Brown, P. D., Schneider, B., Oldham, N. J., Stauber, E., Tokuhisa, J., et al. (2002). Benzoic acid glucosinolate esters and other glucosinolates from *Arabidopsis thaliana*. *Phytochemistry* 59, 663–671. doi: 10.1016/S0031-9422(02)00014-6
- Ruan, Y., Gilmore, J., and Conner, T. (1998). Towards *Arabidopsis* genome analysis: monitoring expression profiles of 1400 genes using cDNA microarrays. *Plant J.* 15, 821–833. doi: 10.1046/j.1365-3113.1998.00254.x
- Shirakawa, M., and Hara-Nishimura, I. (2018). Specialized vacuoles of myrosin cells: chemical defense strategy in Brassicales plants. *Plant Cell Physiol.* 59, 1309–1316. doi: 10.1093/pcp/pcy082
- Shroff, R., Vergara, E., Muck, A., Svatoš, A., and Gershenzon, J. (2008). Nonuniform distribution of glucosinolates in *Arabidopsis thaliana* leaves has important consequences for plant defense. *Proc. Natl. Acad. Sci. USA* 105, 6196–6201. doi: 10.1073/pnas.0711730105
- Sobahan, M. A., Akter, N., Okuma, E., Uraji, M., Ye, W., Mori, I. C., et al. (2015). Allyl isothiocyanate induces stomatal closure in *Vicia faba*. *Biosci. Biotechnol. Biochem.* 79, 1737–1742. doi: 10.1080/09168451.2015.1045827
- Sweetlove, L. J., and Fernie, A. R. (2013). The spatial organization of metabolism within the plant cell. *Annu. Rev. Plant Biol.* 64, 723–746. doi: 10.1146/annurev-arplant-050312-120233
- Taipalensuu, J., Falk, A., and Rask, L. (1996). A wound-and methyl jasmonate-inducible transcript coding for a myrosinase-associated protein with similarities to an early nodulin. *Plant Physiol.* 110, 483–491. doi: 10.1104/pp.110.2.483
- Thangstad, O. P., Gilde, B., Chadchawan, S., Seem, M., Husebye, H., Bradley, D., et al. (2004). Cell specific, cross-species expression of myrosinases in *Brassica napus*, *Arabidopsis thaliana* and *Nicotiana tabacum*. *Plant Mol. Biol.* 54, 597–611. doi: 10.1023/B:PLAN.0000038272.99590.10
- Ueda, H., Nishiyama, C., Shimada, T., Koumoto, Y., Hayashi, Y., Kondo, M., et al. (2006). AtVAM3 is required for normal specification of idioblasts, myrosin cells. *Plant Cell Physiol.* 47, 164–175. doi: 10.1093/pcp/pci232
- Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., et al. (2000). A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403, 623–627. doi: 10.1038/35001009
- Vilakazi, C. S., Dubery, I. A., and Piater, L. A. (2017). Identification of lipopolysaccharide-interacting plasma membrane-type proteins in *Arabidopsis thaliana*. *Plant Physiol. Biochem.* 111, 155–165. doi: 10.1016/j.plaphy.2016.11.025
- Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., et al. (2010). The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* 38(suppl. 2), W214–W220. doi: 10.1093/nar/gkq537
- Wienkoop, S., Zoeller, D., Ebert, B., Simon-Rosin, U., Fisahn, J., Glinski, M., et al. (2004). Cell-specific protein profiling in *Arabidopsis thaliana* trichomes: identification of trichome-located proteins involved in sulfur metabolism and detoxification. *Phytochemistry* 65, 1641–1649. doi: 10.1016/j.phytochem.2004.03.026
- Wittstock, U., and Burow, M. (2010). Glucosinolate breakdown in *Arabidopsis*: mechanism, regulation and biological significance. *Arabidopsis Book* 8:e0134. doi: 10.1199/tab.0134
- Wittstock, U., Kliebenstein, D. J., Lambrix, V., Reichelt, M., and Gershenzon, J. (2003). “Chapter five-Glucosinolate hydrolysis and its impact on generalist and specialist insect herbivores” in *Recent Advances in Phytochemistry*. Vol. 37, ed. J. T. Romeo (London, UK: Elsevier), 101–125.
- Wittstock, U., Kurzbach, E., Herfurth, A.-M., and Stauber, E. J. (2016a). “Chapter six-Glucosinolate breakdown” in *Advances in Botanical Research*. Vol. 80, ed. S. Kopriva (Cambridge, MA, USA: Academic Press), 126–159.
- Wittstock, U., Meier, K., Dörr, F., and Ravindran, B. M. (2016b). NSP-dependent simple nitrile formation dominates upon breakdown of major aliphatic glucosinolates in roots, seeds, and seedlings of *Arabidopsis thaliana* columbia-0. *Front. Plant Sci.* 7:1821. doi: 10.3389/fpls.2016.01821
- Wu, L., Yu, M., Holowachuk, J., Sharpe, A., Lydiate, D., Hegedus, D., et al. (2017). Overexpression of *Brassica napus* myrosinase-associated protein 1 improved *Sclerotinia sclerotiorum* tolerance in *Arabidopsis thaliana*. *Can. J. Plant Sci.* 97, 842–851. doi: 10.1139/cjps-2016-0255
- Xiang, X., Song, M., Zhao, W., Tong, J., Zhang, L., Xiao, L., et al. (2011). A jacalin-related lectin-like gene in wheat is a component of the plant defence system. *J. Exp. Bot.* 10, 1–13. doi: 10.1093/jxb/err226
- Xu, Z., Escamilla-Treviño, L., Zeng, L., Lalgondar, M., Bevan, D., Winkel, B., et al. (2004). Functional genomic analysis of *Arabidopsis thaliana* glycoside hydrolase family 1. *Plant Mol. Biol.* 55, 343–367. doi: 10.1007/s11103-004-0790-1

- Xu, J., Meng, J., Meng, X., Zhao, Y., Liu, J., Sun, T., et al. (2016). Pathogen-responsive MPK3 and MPK6 reprogram the biosynthesis of indole glucosinolates and their derivatives in *Arabidopsis* immunity. *Plant Cell* 28, 1144–1162. doi: 10.1105/tpc.15.00871
- Xue, J., Jørgensen, M., Pihlgren, U., and Rask, L. (1995). The myrosinase gene family in *Arabidopsis thaliana*: gene organization, expression and evolution. *Plant Mol. Biol.* 27, 911–922. doi: 10.1007/BF00037019
- Yan, X., and Chen, S. (2007). Regulation of plant glucosinolate metabolism. *Planta* 226, 1343–1352. doi: 10.1007/s00425-007-0627-7
- Zabala Mde, T., Grant, M., Bones, A. M., Bennett, R., Lim, Y. S., Kissen, R., et al. (2005). Characterisation of recombinant epithiospecifier protein and its over-expression in *Arabidopsis thaliana*. *Phytochemistry* 66, 859–867. doi: 10.1016/j.phytochem.2005.02.026
- Zhang, F., Liu, S., Li, L., Zuo, K., Zhao, L., and Zhang, L. (2016). Genome-wide inference of protein-protein interaction networks identifies crosstalk in abscisic acid signaling. *Plant Physiol.* 171, 1511–1522. doi: 10.1104/pp.16.00057
- Zhang, Z., Ober, J. A., and Kliebenstein, D. J. (2006). The gene controlling the quantitative trait locus EPITHIOSPECIFIER MODIFIER1 alters glucosinolate hydrolysis and insect resistance in *Arabidopsis*. *Plant Cell* 18, 1524–1536. doi: 10.1105/tpc.105.039602
- Zhang, J., Pontoppidan, B., Xue, J., Rask, L., and Meijer, J. (2002). The third myrosinase gene TGG3 in *Arabidopsis thaliana* is a pseudogene specifically expressed in stamen and petal. *Physiol. Plant.* 115, 25–34. doi: 10.1034/j.1399-3054.2002.1150103.x
- Zhao, Y., Hull, A. K., Gupta, N. R., Goss, K. A., Alonso, J., Ecker, J. R., et al. (2002). Trp-dependent auxin biosynthesis in *Arabidopsis*: involvement of cytochrome P450s CYP79B2 and CYP79B3. *Genes Dev.* 16, 3100–3112. doi: 10.1101/gad.1035402
- Zhao, Z., Zhang, W., Stanley, B. A., and Assmann, S. M. (2008). Functional proteomics of *Arabidopsis thaliana* guard cells uncovers new stomatal signaling pathways. *Plant Cell* 20, 3210–3226. doi: 10.1105/tpc.108.063263
- Zhu, M., and Assmann, S. M. (2017). Metabolic signatures in response to abscisic acid (ABA) treatment in *Brassica napus* guard cells revealed by metabolomics. *Sci. Rep.* 7:12875. doi: 10.1038/s41598-017-13166-w
- Zhu, M., Zhu, N., Song, W. Y., Harmon, A. C., Assmann, S. M., and Chen, S. (2014). Thiol-based redox proteins in abscisic acid and methyl jasmonate signaling in *Brassica napus* guard cells. *Plant J.* 78, 491–515. doi: 10.1111/tpj.12490
- Zimmermann, P., Hirsch-Hoffmann, M., Hennig, L., and Gruissem, W. (2004). GENEVESTIGATOR. *Arabidopsis* microarray database and analysis toolbox. *Plant Physiol.* 136, 2621–2632. doi: 10.1104/pp.104.046367

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Chhajed, Misra, Tello and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Evolutionary Developments in Plant Specialized Metabolism, Exemplified by Two Transferase Families

Hiroaki Kusano¹, Hao Li¹, Hiroshi Minami², Yoshihiro Kato², Homare Tabata² and Kazufumi Yazaki^{1*}

OPEN ACCESS

Edited by:

Kazuki Saito,
RIKEN Center for Sustainable
Resource Science (CSRS), Japan

Reviewed by:

Toru Nakayama,
Tohoku University, Japan
Suvi Tuulikki Häkkinen,
VTT Technical Research Centre of
Finland Ltd, Finland

*Correspondence:

Kazufumi Yazaki
yazaki@rsh.kyoto-u.ac.jp

Specialty section:

This article was submitted to
Plant Metabolism and
Chemodiversity,
a section of the journal
Frontiers in Plant Science

Received: 01 March 2019

Accepted: 31 May 2019

Published: 25 June 2019

Citation:

Kusano H, Li H, Minami H, Kato Y,
Tabata H and Yazaki K (2019)
Evolutionary Developments in Plant
Specialized Metabolism, Exemplified
by Two Transferase Families.
Front. Plant Sci. 10:794.
doi: 10.3389/fpls.2019.00794

¹Laboratory of Plant Gene Expression, Research Institute for Sustainable Humansphere, Kyoto University, Kyoto, Japan,
²Life Science Center, Hokkaido Mitsui Chemicals, Sunagawa, Japan

Plant specialized metabolism emerged from the land colonization by ancient plants, becoming diversified along with plant evolution. To date, more than 1 million metabolites have been predicted to exist in the plant kingdom, and their metabolic processes have been revealed on the molecular level. Previous studies have reported that rates of evolution are greater for genes involved in plant specialized metabolism than in primary metabolism. This perspective introduces topics on the enigmatic molecular evolution of some plant specialized metabolic processes. Two transferase families, BAHD acyltransferases and aromatic prenyltransferases, which are involved in the biosynthesis of paclitaxel and meroterpenes, respectively, have shown apparent expansion. The latter family has been shown to be involved in the biosynthesis of a variety of aromatic substances, including prenylated coumarins in citrus plants and shikonin in *Lithospermum erythrorhizon*. These genes have evolved in the development of each special subfamily within the plant lineage. The broadness of substrate specificity and the exon-intron structure of their genes may provide hints to explain the evolutionary process underlying chemodiversity in plants.

Keywords: prenyltransferase, acyltransferase, BAHD, *Citrus*, gene family, molecular evolution, specialized metabolism, *Lithospermum*, *Taxus*

INTRODUCTION

Since land plant colonization 500 million years ago, plant specialized metabolic processes have expanded considerably, resulting in the development of diverse traits within the plant kingdom (Weng et al., 2012). The chemical diversity of those natural products provides various metabolites beneficial for human life, including compounds associated with flavor,

color, taste, and medicine. A comparative genome analysis strongly suggested that gene duplications played a major role in the evolution of divergent metabolic pathways (Fani and Fondi, 2009). The increase in the number of gene copies may have allowed promiscuous diversity of the encoded enzymes, resulting in the synthesis of new metabolites and providing organismal fitness that enhances the establishment of biosynthetic pathways in the plant lineage. The expansion of plant specialized metabolism has been observed in the genome of *Selaginella moellendorffii*, a plant that diverged shortly after the establishment of vascular tissues in plant evolution (Banks et al., 2011). A representative example of these expanded gene families is cytochrome P450-dependent monooxygenases, which constitute 1% of the predicted proteome in *Selaginella*. The genome of liverwort, *Marchantia polymorpha*, also encodes many terpenoid biosynthetic enzymes sharing a common isoprenoid pathway, a derivative designated taxadiene for the synthesis of plant hormones like gibberellin (Bowman et al., 2017). In *Physcomitrella patens*, a diterpene *ent*-kaurene is converted to gibberellin-type diterpenes, which act as regulators of protonema differentiation (Hayashi et al., 2010).

Species of the gymnosperm *Taxus* synthesize unique diterpene compounds called “taxoids,” which include an important anticancer drug, paclitaxel, a derivative designated taxadiene (Guerra-Bubb et al., 2012). Over 350 taxoid compounds were identified by 1999, with these compounds having variable side residues at the C1, C2, C4, C5, C7, C9, C10, C13, and C14 positions of the core taxadiene skeleton (Baloglu and Kingston, 1999). Except for a partial biosynthetic route (Croteau et al., 2006), knowledge about the biosynthetic pathway of taxoids that contribute to the chemodiversity in *Taxus* is limited.

Because of their fine-tuned genome data resources, angiosperm species provide good model systems to study molecular mechanisms underlying the chemodiversity of plant metabolites (Kroymann, 2011). For example, meroterpenes, including furanocoumarin derivatives (Bourgaud et al., 2006) and shikonin derivatives that are lipophilic red naphthoquinone (Yazaki, 2017), are specialized metabolites synthesized through branched routes from a metabolic pathway common to the general phenylpropanoid and isoprenoid biosynthetic pathways (Yazaki et al., 2017). The term “primary metabolism” indicates processes required to sustain life, such as energy acquisition from glucose. These processes include, for example, the biosynthesis of ubiquinone, a component of the respiratory chain in mitochondria. The biosynthesis of shikonin derivatives involves steps common to those involved in ubiquinone biosynthesis. To avoid confusion in distinguishing between primary and specialized (secondary) metabolism, this article uses the term “common metabolism” rather than “primary metabolism” to indicate biosynthetic pathways conserved in a broad variety of organisms.

This perspective focuses on two enzyme families as examples of molecular evolutionary events: the aromatic substrate prenyltransferase family, which plays a key role in the diversification of phenolics, and the BAHD (BEAT-AHCT-HCBT-DAT; initials of representative members) acyltransferase family, which is responsible for the derivatization of a core metabolite.

EVOLUTION OF THE CITRUS PRENYLTRANSFERASE GENE FAMILY

Among prenyltransferase superfamily including prenyl chain elongation enzymes, aromatic prenyltransferases represent a family responsible for the prenylation of aromatic substances. An aromatic prenyltransferase of *Citrus limon*, CIPT1, is responsible for the biosynthesis of 8-geranylumbelliferone, a coumarin derivative of a plant specialized metabolite (Munakata et al., 2014). The chemical diversity of coumarin derivatives is greatly increased by the involvement of aromatic prenyltransferases, which have been identified in many plant lineages during the last decade (Karamat et al., 2014; Munakata et al., 2014). Phylogenetic analysis has suggested that the diverse prenyltransferases developed independently in each plant family rather than developing from a common ancestor within the prenyltransferase gene family (Munakata et al., 2016). The plant prenyltransferase gene family contains conserved subfamilies responsible for the ubiquinone, plastoquinone, and vitamin E biosynthesis pathways (Li, 2016).

An outline of the evolutionary development of plant aromatic prenyltransferases in *Citrus* species was revealed by a phylogenetic analysis of previously characterized prenyltransferases and prenyltransferases of the model species *P. patens*, *S. moellendorffii*, *Arabidopsis thaliana*, *Glycine max*, and *Lithospermum erythrorhizon* (see below), in addition to *Citrus sinensis* (Figure 1A). Phylogenetically, these intrinsic membrane proteins can be grouped into three major subfamilies, i.e., those involved in the biosynthesis of vitamin E, plastoquinone, and ubiquinone (shown as yellow and gray backgrounds and as the black triangle, respectively in Figure 1A, with the black triangle expanded in Figure 1B). The biochemical functions of AtVTE2-1 (Savidge et al., 2002), AtVTE2-2 (Venkatesh et al., 2006), and OsPPT1 (Ohara et al., 2006) have been described. As expected from their fundamental roles, all model plant species had one or more proteins in each subfamily. In contrast, a search of the *C. sinensis* database revealed nine prenyltransferase-like proteins, forming a *Citrus*-specific subfamily within the vitamin E clade (shown in red in Figure 1A). A similar result was obtained by searching *Citrus clementina* genome sequences. These results suggest that *Citrus* species have developed a unique, expanded gene subfamily for specialized metabolism, with CIPT1 being biochemically characterized. This analysis also identified a similar unique subfamily expansion in *G. max* (shown in blue in Figure 1A). The first flavonoid-specific prenyltransferase SfN8DT1 from a legume species *Sophora flavescens* (Sasaki et al., 2008) is in this group, suggesting that flavonoid prenyltransferases in soybeans were derived from a vitamin E biosynthetic enzyme. Other, later detected flavonoid prenyltransferases were all classified in this subgroup (Akashi et al., 2009; Yoneyama et al., 2016). Most of these enzymes involved in specialized metabolism show strict substrate specificity in relation to a particular prenyl diphosphate.

Prenyltransferases involved in common metabolism show broad specificity in relation to substrates of different side chain lengths; i.e., they accept various prenyl diphosphates of different chain lengths (Sadre et al., 2010). For example, the ubiquinone biosynthesis pathway in rice can be modified by introducing a decaprenyl diphosphate synthase, resulting in the production of non-native UQ10 rather than native UQ9 (Ohara et al., 2006; Takahashi

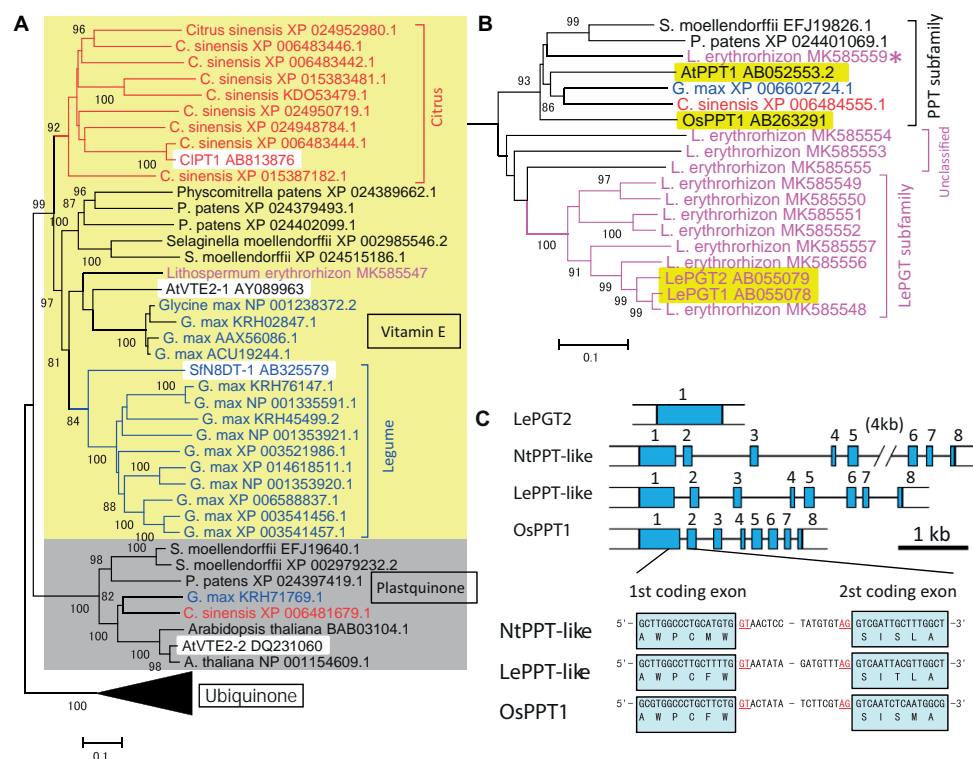


FIGURE 1 | Phylogenetic analysis of the prenyltransferase family in citrus species and *Lithospermum erythrorhizon* as well as in model plants. **(A)** Grouping of plant prenyltransferases into three major clades: a clade represented by the *Arabidopsis* homogentisate phytyltransferase AtVTE2-1 involved in vitamin E biosynthesis (indicated by “Vitamin E” and a yellow background), a clade represented by *Arabidopsis* AtVTE2-2 for plastoquinone biosynthesis (indicated by “Plastoquinone” and a gray background), and a clade represented by the rice polyprenyltransferase OsPPT1 for ubiquinone biosynthesis (indicated by “Ubiquinone” and a compressed black triangle). Biochemically characterized proteins are indicated by a white background. The *Citrus* and legume proteins are shown in red and blue letters, respectively, and the lineage-specific clades are indicated by brackets with the same colors. **(B)** Details of the phylogenetic tree of polyprenyltransferases for ubiquinone in panel (A). *L. erythrorhizon* proteins are shown in magenta letters. The brackets indicate subclades of polyprenyltransferases involved in ubiquinone biosynthesis (PPT subfamily), LePGT-like proteins (LePGT subfamily), and unclassified subclade proteins (unclassified). The proteins from other organisms are shown in black letters. The asterisk indicates the putative PPT-like protein of *L. erythrorhizon*. The phylogenetic tree was drawn using the MEGA7 neighbor-joining method with 1,000 bootstrap replicates for alignment of polyprenyltransferase-related proteins, which were calculated with the MUSCLE algorithm. The accession numbers are shown next to the name of the organism. Biochemically characterized proteins are indicated by a yellow background. The scale bar represents 0.1 amino acid substitutions per site. **(C)** LePGT gene is encoded by a single exon gene in the *L. erythrorhizon* genome, whereas LePPT-like proteins are encoded by genes with inserted introns, at positions similar to those of the authorized OsPPT gene and the closest tobacco homolog, NtPPT-like gene (gene = LOC107804153). The first intron insertion site into the coding region is shown. Scale bar, 1 kb of DNA sequence. Blue boxes represent coding exons.

et al., 2006). These expanded gene families and the broad substrate specificity of prenyltransferases may provide the opportunity for neo-functionalization of new enzymes in plant evolutionary history.

EVOLUTION OF THE *P*-HYDROXYBENZOIC ACID GERANYLTRANSFERASE GENE FOR SHIKONIN BIOSYNTHESIS

A boraginaceae medicinal plant, *L. erythrorhizon*, possesses a unique subfamily of *p*-hydroxybenzoic acid geranyltransferases (PGTs) (Figure 1B) that are specifically involved in shikonin biosynthesis (Yazaki et al., 2002). An overview of the evolutionary history of PGT was attained by assessing genome sequences and transcriptomes of *L. erythrorhizon* from the GenBank datasets SRP108575 and SRP141330, respectively, as well as by reassembling our original data (Takanashi et al., 2019). The

hypothetical PGT-like proteins were found to be closely related to the ubiquinone prenyltransferase subfamily involved in common metabolism (magenta in Figure 1B), which was closer to these hypothetical PGT-like proteins than the specialized citrus prenyltransferases (Figure 1A). Most PGT-like proteins are encoded by genes with a single exon, whereas general ubiquinone biosynthetic polyprenyltransferases (PPTs) are encoded by genes containing multiple exons (Figure 1C). It is of interest to determine how the single exon structure was generated during the evolution of plant specialized metabolism.

MISSING UBIQUINONE PRENYLTRANSFERASE IN *L. ERYTHRORHIZON*

Although ubiquinone is a common metabolite in all eukaryotes, and the genes encoding PPTs are essential for the survival of

a broad range of organisms, no orthologous ubiquinone PPT gene was found in the *L. erythrorhizon* transcriptome. Experiments in yeast showed that LePGT cannot synthesize ubiquinone (Yazaki et al., 2002), and LePGT heterologously expressed in *E. coli* was found to inhibit ubiquinone biosynthesis (Wu et al., 2015). Genomic sequencing identified a contig fragment that could code for PPT-like proteins (asterisk in **Figure 1B**) and that its amino acid sequence was moderately similar to that of OsPPT1, which is responsible for ubiquinone biosynthesis in rice. In addition, there were three contigs that we could not classify, which are labeled “unclassified genes” (“unclassified” in **Figure 1B**). In contrast to the particular PGT that catalyzes shikonin biosynthesis, an intron insertion was found in the hypothetical gene, at the same position as in the PGTs of *Nicotiana tabacum* and *Oryza sativa* (**Figure 1C**). This conserved exon-intron organization was also observed in the PPT genes from *Arabidopsis* and rice (Ohara et al., 2006). This gene product is a strong candidate for a ubiquinone prenyltransferase in *L. erythrorhizon*, and its biochemical characterization is expected in the future.

EVOLUTION OF THE *TAXUS* ACYLTRANSFERASE GENE FAMILY

Acyltransferases also substantially contribute to the diversification of specialized metabolites, in which BAHD and SCPL (serine carboxypeptidase-like) are representatives. Taxoids such as paclitaxel present in *Taxus* species are specialized metabolites and highly acylated compounds. Five known taxoid acyltransferases are closely related to each other, with all grouped in clade V of the BAHD acyltransferase family (D'Auria, 2006). These *Taxus* proteins differ in substrate specificities for both acyl donors and acceptors; i.e., they can utilize acetyl-CoA, benzoyl-CoA or phenylalanoyl-CoA for O- and N-acylation of various taxoid molecules (D'Auria, 2006).

To understand the evolutionary development of the *Taxus* BAHD acyltransferase family, BAHD clade V was analyzed phylogenetically in detail (yellow background in **Figure 2A**). The amino acid sequences of *Taxus* BAHD members were obtained from the transcriptome data of *Taxus x media* cultured cells (Yukimune et al., 1996). Phylogenetic analysis showed that the *Taxus* BAHD proteins form a *Taxus*-specific clade (red bracket in **Figure 2A**), containing all five characterized acyltransferases (white background in the *Taxus*-specific clade), as well as other *Taxus* proteins of unknown function (asterisk in **Figure 2A**). Within this clade of the BAHD family, *O. sativa* and *A. thaliana* each form a unique clade, suggesting that lineage-specific subfamily expansion of the BAHD acyltransferases plays a major role in plant evolution (Fani and Fondi, 2009). In addition to this *Taxus*-specific subgroup, other *Taxus* BAHD proteins have been identified, with these classified with other model plant BAHD members (**Supplementary Figure S1**), suggesting that *Taxus* species possess genes encoding general BAHD clade V proteins that are conserved among a broad range of plant species.

It can be hypothesized that neo-functionalization is induced by the acquisition of promiscuous enzymatic activity during

plant evolution. We have examined the enzymatic activity of recombinant proteins prepared from seven isolated cDNAs encoding BAHD members of the *Taxus*-specific subfamily (dagger in **Figure 2A**). Each crude recombinant enzyme was prepared using pET22a and OrigamiB as a host-vector system (Novagen), without a periplasmic signal sequence, according to the conventional method. Each enzyme was reacted with acetyl-CoA and 10-deacetyl baccatin III (10-DAB) as substrates, and the reaction products were analyzed using an UPLC-MS/MS system equipped with a BEH C18 column (Waters). The clone encoding 5-hydroxytaxadiene 5-O-acetyltransferase (TAT) had 10-DAB:10-O-acetyltransferase (DBAT) activity (Walker et al., 2000), as well as the canonical enzyme DBAT (**Figures 2B,C**; Walker and Croteau, 2000). The amount of the product formed by the substrate was 1.4 mol% for TAT and 10.4% for DBAT, suggesting that the activity of TAT was 13.2% that of DBAT. This promiscuity of enzymatic activity may represent the evolutionary footprint of a biosynthetic enzyme that acquires a new functionality through the alteration of substrate and product specificities, resulting in the production of a unique specialized metabolite.

CONCLUSIONS AND PERSPECTIVES

Using two transferase subfamilies as examples, we have shown the “heritage” of expansion of a gene family, which is relevant for the development of plant specialized metabolic pathways. A protein in the specific BAHD subfamily of *Taxus* species showed promiscuous enzymatic activity for noncanonical substrates containing side chains at a noncanonical carbon position. These observations fit the general context of developmental molecular evolution that explains the development and establishment of new canonical enzymatic activity (Weng et al., 2012). The generation in *L. erythrorhizon* of a PGT gene subfamily, each containing a single exon and involved in shikonin biosynthesis, suggests the putative involvement of the reverse transcription of mature mRNA. If this surmise is valid for other enzyme families, single exon genes may provide clues to identifying missing proteins responsible for biosynthetic pathways for other valuable plant specialized metabolites.

There are yet many missing links, even in actively studied shikonin and taxoid biosynthetic pathways. The applicable range of the single exon hypothesis may not be limited only to biosynthetic enzymes, but to regulatory factors. The identification of regulatory factors will be essential to understanding the production of plant specialized metabolites, including membrane transporters. Comparative genomics will enable the assessment of the evolutionary footprint of these genes, e.g., the expansion of specific subfamilies and the proliferation of single exon genes. Further biochemical and molecular genetics studies may provide experimental evidence for the involvement of hypothetical proteins in plant specialized metabolism.

DATA AVAILABILITY

The datasets generated for this study can be found in GenBank.

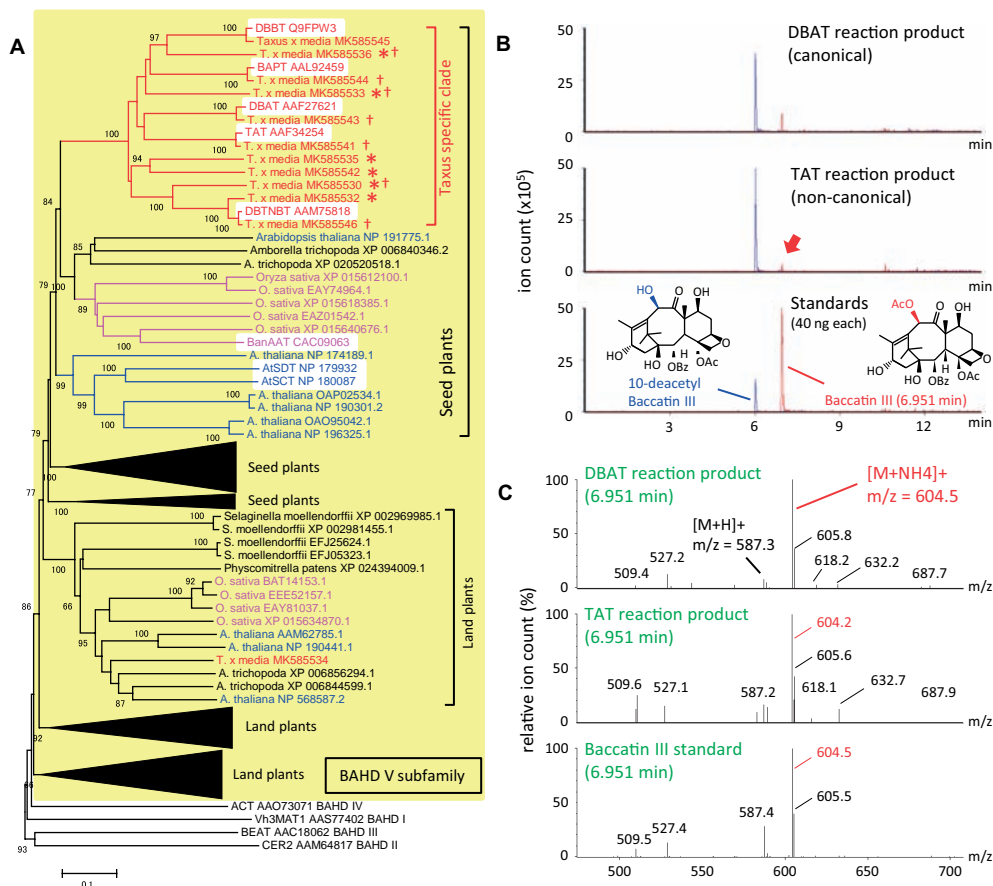


FIGURE 2 | Phylogenetic analysis of BAHD acyltransferase proteins from *Taxus* species and LC-MS/MS analysis of the reaction products of the noncanonical enzyme, taxadienol 5-acyltransferase. **(A)** Performance of phylogenetic analysis with hypothetical *Taxus* BAHD acyltransferase-like proteins and related proteins from model plant species. The BAHD family was classified into five clades (D'Auria, 2006), with clade V indicated by a yellow background, and representatives of clade I–IV (Vh3MAT1, CER2, BEAT, and ACT, respectively) placed outside the yellow background. Proteins of *Taxus*, rice, *Arabidopsis* are shown in red, magenta, and blue letters, respectively, and the lineage-specific subclades are indicated by the same colors. The bracket “*Taxus* specific clade” indicates the *Taxus* lineage-specific subclade containing the five characterized proteins, TAT, DBAT, DBTNBT, DBBT, and BAPT, indicated by a white background. Asterisks indicate *Taxus* proteins of unknown function, and daggers indicate proteins biochemically analyzed in the present study. A representative widely conserved clade in land plants from *Physcomitrella* to *Arabidopsis* is indicated by brackets, with four other subclades compressed (expanded in **Supplementary Figure S1**), in addition to the clade conserved in seed plants containing the *Taxus* specific clade. The accession numbers are given next to the organism names. The phylogenetic tree was drawn using the MEGA7 neighbor-joining method with 1,000 bootstrap replicates for alignment calculated with the MUSCLE algorithm. Scale bar, 0.1 amino acid substitutions per site. **(B)** LC-MS/MS chromatograms of the enzyme reaction products of *Taxus* acyltransferases DBAT and TAT using acetyl-CoA and 10-DAB as substrates. The red arrow indicates the peak of the noncanonical reaction product. The bottom panel shows the chromatogram of standard specimens, 10-DAB and baccatin III. The chromatograms show a trace of representative ions $m/z = 545.5$ $[M + H]^+$ and 604.5 $[M + NH_4]^+$ for the substrate 10-DAB (blue) and the product baccatin III (red), respectively. The vertical axis indicates the value relative to 5 million ion counts. **(C)** Mass spectrum of the *in vitro* reaction product peaks found at a retention time of 6.951 min of the chromatogram. The vertical axis indicates the relative value of ion count of maximum signal at $m/z = 604.5$. The molecular formulas of 10-DAB and baccatin III are shown in panel **(B)**.

AUTHOR CONTRIBUTIONS

HK and KY wrote the manuscript and performed the phylogenetic and biochemical analyses. HL was involved in the assembly of genomic contigs and the analysis of the exon-intron structure of *Lithospermum erythrorhizon* genes. HM, YK, and HT were responsible for transcriptome analysis of *Taxus* spp.

FUNDING

This work was supported in part by the New Energy and Industrial Technology Development Organization (NEDO, No.

16100890 to KY). Additional support was provided by the Mission Research of RISH, Kyoto University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.00794/full#supplementary-material>

FIGURE S1 | Expanded phylogenetic tree of **Figure 2A**. Phylogenetic analysis with hypothetical *Taxus* BAHD acyltransferase-like proteins and related proteins from model plant species. Asterisks indicate *Taxus* proteins found in this study. Functionally identified BAHD proteins are highlighted in yellow background.

REFERENCES

- Akashi, T., Sasaki, K., Aoki, T., Ayabe, S., and Yazaki, K. (2009). Molecular cloning and characterization of a cDNA for pterocarpan 4-dimethylallyltransferase catalyzing the key prenylation step in the biosynthesis of glyceollin, a soybean phytoalexin. *Plant Physiol.* 149, 683–693. doi: 10.1104/pp.108.123679
- Baloglu, E., and Kingston, D. G. I. (1999). The taxane diterpenoids. *J. Nat. Prod.* 62, 1448–1472. doi: 10.1021/np990176i
- Banks, J. A., Nishiyama, T., Hasebe, M., Bowman, J. L., Gribskov, M., dePamphilis, C., et al. (2011). The Selaginella genome identifies genetic changes associated with the evolution of vascular plants. *Science* 332, 960–963. doi: 10.1126/science.1203810
- Bourgaud, F., Hehn, A., Larbat, R., Doerper, S., Gontier, E., Kellner, S., et al. (2006). Biosynthesis of coumarins in plants: a major pathway still to be unravelled for cytochrome P450 enzymes. *Phytochem. Rev.* 5, 293–308. doi: 10.1007/s11101-006-9040-2
- Bowman, J. L., Kohchi, T., Yamato, K. T., Jenkins, J., Shu, S., Ishizaki, K., et al. (2017). Insights into land plant evolution garnered from the Marchantia polymorpha genome. *Cell* 171, 287–304. doi: 10.1016/j.cell.2017.09.030
- Croteau, R., Ketchum, R. E. B., Long, R. M., Kaspera, R., and Wildung, M. R. (2006). Taxol biosynthesis and molecular genetics. *Phytochem. Rev.* 5, 75–97. doi: 10.1007/s11101-005-3748-2
- D'Auria, J. C. (2006). Acyltransferases in plants: a good time to be BAHD. *Curr. Opin. Plant Biol.* 9, 331–340. doi: 10.1016/j.pbi.2006.03.016
- Fani, R., and Fondi, M. (2009). Origin and evolution of metabolic pathways. *Phys. Life. Rev.* 6, 23–52. doi: 10.1016/j.plrev.2008.12.003
- Guerra-Bubb, J., Croteau, R., and Williams, R. M. (2012). The early stages of taxol biosynthesis: an interim report on the synthesis and identification of early pathway metabolites. *Nat. Prod. Rep.* 29, 683–696. doi: 10.1039/c2np20021j
- Hayashi, K., Horie, K., Hiwatashi, Y., Kawaide, H., Yamaguchi, S., Hanada, A., et al. (2010). Endogenous diterpenes derived from ent-kaurene, a common gibberellin precursor, regulate protonema differentiation of the moss *Physcomitrella patens*. *Plant Physiol.* 153, 1085–1097. doi: 10.1104/pp.110.157909
- Karamat, F., Olry, A., Munakata, R., Koeduka, T., Sugiyama, A., Paris, C., et al. (2014). A coumarin-specific prenyltransferase catalyzes the crucial biosynthetic reaction for furanocoumarin formation in parsley. *Plant J.* 77, 627–638. doi: 10.1111/tpj.12409
- Kroymann, J. (2011). Natural diversity and adaptation in plant secondary metabolism. *Curr. Opin. Plant Biol.* 14, 246–251. doi: 10.1016/j.pbi.2011.03.021
- Li, W. (2016). Bringing bioactive compounds into membranes: the UbiA superfamily of intramembrane aromatic prenyltransferases. *Trends Biochem. Sci.* 41, 356–370. doi: 10.1016/j.tibs.2016.01.007
- Munakata, R., Inoue, T., Koeduka, T., Karamat, F., Olry, A., Sugiyama, A., et al. (2014). Molecular cloning and characterization of a geranyl diphosphate-specific aromatic prenyltransferase from lemon. *Plant Physiol.* 166, 80–90. doi: 10.1104/pp.114.246892
- Munakata, R., Olry, A., Karamat, F., Courdavault, V., Sugiyama, A., Date, Y., et al. (2016). Molecular evolution of parsnip (*Pastinaca sativa*) membrane-bound prenyltransferases for linear and/or angular furanocoumarin biosynthesis. *New Phytol.* 211, 332–344. doi: 10.1111/nph.13899
- Ohara, K., Yamamoto, K., Hamamoto, M., Sasaki, K., and Yazaki, K. (2006). Functional characterization of OsPPT1, which encodes p-hydroxybenzoate polyprenyltransferase involved in ubiquinone biosynthesis in *Oryza sativa*. *Plant Cell Physiol.* 47, 581–590. doi: 10.1093/pcp/pcj025
- Sadre, R., Frentzen, M., Saeed, M., and Hawkes, T. (2010). Catalytic reactions of the homogentisate prenyl transferase involved in plastoquinone-9 biosynthesis. *J. Biol. Chem.* 285, 18191–18198. doi: 10.1074/jbc.M110.117929
- Sasaki, K., Mito, K., Ohara, K., Yamamoto, H., and Yazaki, K. (2008). Cloning and characterization of naringenin 8-prenyltransferase, a flavonoid-specific prenyltransferase of *Sophora flavescens*. *Plant Physiol.* 146, 1075–1084. doi: 10.1104/pp.107.110544
- Savidge, B., Weiss, J. D., Wong, Y. H., Lassner, M. W., Mitsky, T. A., Shewmaker, C. K., et al. (2002). Isolation and characterization of homogentisate phytyltransferase genes from *Synechocystis* sp. PCC 6803 and *Arabidopsis*. *Plant Physiol.* 129, 321–332. doi: 10.1104/pp.010747
- Takahashi, S., Ogiyama, Y., Kusano, H., Shimada, H., Kawamukai, M., and Kadowaki, K. (2006). Metabolic engineering of coenzyme Q by modification of isoprenoid side chain in plant. *FEBS Lett.* 580, 955–959. doi: 10.1016/j.febslet.2006.01.023
- Takanashi, K., Nakagawa, Y., Aburaya, S., Kaminade, K., Aoki, W., Saida-Munakata, Y., et al. (2019). Comparative proteomic analysis of *Lithospermum erythrorhizon* reveals regulation of a variety of metabolic enzymes leading to comprehensive understanding of the shikonin biosynthetic pathway. *Plant Cell Physiol.* 60, 19–28. doi: 10.1093/pcp/pcy183
- Venkatesh, T. V., Karunanandaa, B., Free, D. L., Rottnek, J. M., Baszis, S. R., and Valentin, H. E. (2006). Identification and characterization of an *Arabidopsis* homogentisate phytyltransferase paralog. *Planta* 223, 1134–1144. doi: 10.1007/s00425-005-0180-1
- Walker, K., and Croteau, R. (2000). Molecular cloning of a 10-deacetylbaaccatin III-10-O-acetyl transferase cDNA from *Taxus* and functional expression in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 97, 583–587. doi: 10.1073/pnas.97.2.583
- Walker, K., Schoendorf, A., and Croteau, R. (2000). Molecular cloning of a taxa-4(20),11(12)-dien-5 α -ol-O-acetyl transferase cDNA from *Taxus* and functional expression in *Escherichia coli*. *Arch. Biochem. Biophys.* 374, 371–380. doi: 10.1006/abbi.1999.1609
- Weng, J. K., Philippe, R. N., and Noel, J. P. (2012). The rise of chemodiversity in plants. *Science* 336, 1667–1670. doi: 10.1126/science.1217411
- Wu, H., Bennett, G. N., and San, K.-Y. (2015). Metabolic control of respiratory levels in coenzyme Q biosynthesis-deficient *Escherichia coli* strains leading to fine-tune aerobic lactate fermentation. *Biotechnol. Bioeng.* 112, 1720–1726. doi: 10.1002/bit.25585
- Yazaki, K. (2017). *Lithospermum erythrorhizon* cell cultures: present and future aspects. *Plant Biotechnol.* 34, 131–142. doi: 10.5511/plantbiotechnology.17.0823a
- Yazaki, K., Arimura, G., and Ohnishi, T. (2017). “Hidden” terpenoids in plants: their biosynthesis, localization and ecological roles. *Plant Cell Physiol.* 58, 1615–1621. doi: 10.1093/pcp/pcx123
- Yazaki, K., Kuniyama, M., Fujisaki, T., and Sato, F. (2002). Geranyl diphosphate: 4-hydroxybenzoate geranyltransferase from *Lithospermum erythrorhizon*. Cloning and characterization of a key enzyme in shikonin biosynthesis. *J. Biol. Chem.* 277, 6240–6246. doi: 10.1074/jbc.M106387200
- Yoneyama, K., Akashi, T., and Aoki, T. (2016). Molecular characterization of soybean pterocarpan 2-dimethylallyltransferase in glyceollin biosynthesis: local gene and whole-genome duplications of prenyltransferase genes led to the structural diversity of soybean prenylated isoflavonoids. *Plant Cell Physiol.* 57, 2497–2509. doi: 10.1093/pcp/pcw178
- Yukimune, Y., Tabata, H., Higashi, Y., and Hara, Y. (1996). Methyl jasmonate-induced overproduction of paclitaxel and baccatin III in *Taxus* cell suspension cultures. *Nat. Biotechnol.* 14, 1129–1132. doi: 10.1038/nbt0996-1129

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Kusano, Li, Minami, Kato, Tabata and Yazaki. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Acceleration of Mechanistic Investigation of Plant Secondary Metabolism Based on Computational Chemistry

Hajime Sato^{1,2}, Kazuki Saito^{1,3} and Mami Yamazaki^{1*}

¹Graduate School of Pharmaceutical Sciences, Chiba University, Chiba, Japan, ²Center for Sustainable Resource Science, Advanced Elements Chemistry Laboratory, Cluster for Pioneering Research (CPR), RIKEN, Saitama, Japan, ³RIKEN Center for Sustainable Resource Science, Yokohama, Japan

OPEN ACCESS

Edited by:

Hiroyuki Morita,
University of Toyama, Takaoka, Japan

Reviewed by:

Takashi Matsui,
Kitasato University, Japan
Takahiro Mori,
The University of Tokyo, Japan

*Correspondence:

Mami Yamazaki
mamiy@faculty.chiba-u.jp

Specialty section:

This article was submitted to
Plant Metabolism and
Chemodiversity,
a section of the journal
Frontiers in Plant Science

Received: 09 May 2019

Accepted: 04 June 2019

Published: 26 June 2019

Citation:

Sato H, Saito K and Yamazaki M
(2019) Acceleration of Mechanistic
Investigation of Plant Secondary
Metabolism Based on
Computational Chemistry.
Front. Plant Sci. 10:802.
doi: 10.3389/fpls.2019.00802

This review describes the application of computational chemistry to plant secondary metabolism, focusing on the biosynthetic mechanisms of terpene/terpenoid, alkaloid, flavonoid, and lignin as representative examples. Through these biosynthetic studies, we exhibit several computational methods, including density functional theory (DFT) calculations, theozyme calculation, docking simulation, molecular dynamics (MD) simulation, and quantum mechanics/molecular mechanics (QM/MM) calculation. This review demonstrates how modern computational chemistry can be employed as an effective tool for revealing biosynthetic mechanisms and the potential of computational chemistry—for example, elucidating how enzymes regulate regio- and stereoselectivity, finding the key catalytic residue of an enzyme, and assessing the viability of hypothetical pathways. Furthermore, insights for the next research objective involving application of computational chemistry to plant secondary metabolism are provided herein. This review will be helpful for plant scientists who are not well versed with computational chemistry.

Keywords: biosynthesis, computational chemistry, density functional theory, molecular dynamics simulation, quantum mechanics/molecular mechanics, plant, secondary metabolite

INTRODUCTION

Recent advances in life science have revealed many plant secondary biosynthetic pathways and have also contributed to establishing efficient microbial-based manufacturing systems that can potentially afford greater amounts of important plant secondary metabolites, including artemisinin (Ro et al., 2006; Paddon et al., 2013), opioids (Galanie et al., 2015), and cannabinoids (Luo et al., 2019), by introducing all of the biosynthetic genes into the host. These outstanding achievements are obviously based on long-standing biosynthetic studies. In opioid heterologous expression systems, not only enzymes from plants but also enzymes from other organisms were used, which are relevant to the expression levels and substrate specificity. Considering the progress of biosynthetic studies and synthetic biology, rational engineering of the existing biosynthetic pathways or designing the novel pathways to obtain desired products appears to be the next challenge. To achieve this objective, detailed mechanistic investigations of biosynthetic reactions are critical for the rational modification of biosynthetic pathways or enzymes.

Although various types of experimental methods, such as labeling experiments, X-ray crystallography, site-directed mutagenesis, omics studies (Rai et al., 2017, 2019), and genome editing, have been applied, clarifying entire biosynthetic pathways and mechanisms remains a challenge in plant science. Unlike microorganisms, biosynthetic genes do not form clusters in plants, which makes biosynthetic studies of plants difficult. Even finding a biosynthetic gene (enzyme) is a challenge. For example, thus far, in quinolizidine alkaloid biosynthesis, only two enzymes have been characterized for their skeletal formation, although studies on their biosynthesis have been carried out for over half a century (Bunsupa et al., 2012; Yang et al., 2017).

Recently, an increasing number of computational chemistry studies on natural product biosynthesis have been published, which provides new insights that cannot be discovered solely by experimental approach. However, among plant biologists and computational chemists, there is still a significant gap in understanding how computational chemistry is an effectual tool. Through several representative biosynthetic studies on terpene/terpenoid, alkaloid, flavonoid, and lignin, this review describes what can be clarified using computational chemistry and also describes the computational methods, which are widely used for the mechanistic investigation of biosynthetic reactions. An attempt to be comprehensive was made; however, apologies are offered in advance for any accidental oversights. Moreover, there are many computational predictions of NMR, UV, and CD spectra, and mechanistic investigations on fungal or bacterial secondary metabolism; however, these studies are not covered here since this review focuses on the “mechanistic investigations” of “plant” secondary metabolism.

THEORETICAL METHODS FOR NATURAL PRODUCT BIOSYNTHESIS RESEARCH

In plant science, cheminformatics such as KEGG (Kanehisa and Goto, 2000; Kanehisa et al., 2017, 2019) KNApSACk (Afendi et al., 2011), etc., are widely utilized to predict biosynthetic pathways. However, computational chemistry is completely different from these database search approaches. In computational chemistry, the properties of compounds, such as free energy, spectrum, reactivity, and molecular orbitals, etc., are obtained by *ab initio* calculations based on the Schrödinger equation in molecular orbital (MO) calculations, the Kohn-Sham equation in density functional theory (DFT) calculations, and the Newton equation in molecular dynamics (MD) simulations (Jensen, 2017).

Although a variety of computational methods are available for the mechanistic investigation of chemical reactions, carefully choosing appropriate methods is important in terms of accuracy and computational expenses. Today, quantum mechanics (QM), i.e., MO or DFT, are mainly used for the evaluation of the reactivity or properties of small molecules. For mechanistic investigations using QM calculations, transition state (TS) search is initially performed, after which frequency calculation is performed to ensure that the TS has a single imaginary frequency.

Finally, intrinsic reaction coordinate (IRC) calculation is carried out to obtain the reactant and product (Ishida et al., 1977; Fukui, 1981; Gonzalez and Schlegel, 1989; Page et al., 1990; Schlegel and Gonzalez, 1990). Today, hundreds of levels of theory are available. The choice of the level of theory is quite critical for computation accuracy in QM, which is dependent on the type of reactions or chemical structures. Many publications on benchmark tests exist, in which several combinations of density functional and basis set are tested against the one certain reaction or molecule to find the most accurate level of theory. For example, mPW1PW91/6-31+G(d,p)//B3LYP/6-31+G(d,p) has been used for the terpene-forming reaction based on the benchmark test reported by Matsuda et al. (2006). In this literature, many combinations of basis set and functionals were tested and compared with the experimental data. The results indicated that B3LYP/6-31+G(d,p) is the best for geometry optimization, whereas mPW1PW91/6-31+G(d,p) is the best for computing the free energy for the terpene-forming reaction.

In plant secondary metabolism, most of the biosynthetic reactions are thought to be catalyzed by enzymes; however, the system size, which QM calculations can treat, is usually up to a few hundred atoms. This means that biological macromolecules, i.e., enzymes, are too large to be calculated with QM. Theozyme calculation (Tantillo et al., 1998; Ujaque et al., 2002; Tantillo, 2010) is one way to estimate the enzymatic assistance toward the chemical conversion, in which the catalytic center, substrate, and several residues are picked up and subsequently subjected to DFT calculations (see Section “Theozyme Calculation Identified the Key Residue for Sesterfisherol Biosynthesis” for more detail). However, QM calculation is applied only for the isolated model, and the regions that could affect the chemical conversion are ignored. Thus, MD simulation is used for large macromolecule systems, which can simulate the time-dependent structure changes of enzymes; however, the free energy value is less accurate than that obtained using QM. For relatively high accuracy, quantum mechanics/molecular mechanics (QM/MM) is used for large macromolecule systems, in which the reaction system is divided into two regions: QM and MM. Generally, the catalytic center is calculated using QM, and the other parts of the enzyme are calculated using MM. Moreover, a state-of-the-art QM/MM MD is also utilized for mechanistic investigations.

TERPENOIDS

Terpene/terpenoids are the largest natural product group. At least 80,000 terpene/terpenoids have been reported to date (Quin et al., 2014; Dickschat, 2016; Christianson, 2017). Most theoretical studies on terpene cyclization have been carried out by Tantillo and Hong. Although they have provided valuable insight into carbocation chemistry, only some of their works are presented in this review due to page limitations. One of the reasons for the considerable success of this compound group in computational chemistry is that terpenes consist of only carbons and hydrogens; therefore, strong interactions between the substrate and enzyme, such as hydrogen bonding, are not necessarily considered for its cyclization. In fact, the

computed inherent reactivity shows good agreement with the experimental data for the terpene-forming reaction. Moreover, as was described above, a detailed benchmark test has been reported, which also supports the computation accuracy. In this section, several examples ranging from small systems to large systems will be discussed.

Two Possible Cyclization Mechanisms Were Assessed Using Density Functional Theory Calculations

In comparison to mono-, sesqui-, di-, and triterpene, sesterterpenes are considered relatively minor terpenes, and only several biosynthetic enzymes have been characterized to date. In 2017, a genome mining approach to discovering the biosynthetic genes of sesterterpenes from *Arabidopsis thaliana*, *Capsella rubella*, *Brassica oleracea*, and *Brassica rapa* was reported by Huang et al. (2017), which strongly promoted the biosynthetic research of sesterterpenoids. They isolated (+)-arathanatriene, (–)-retigeranin B, (+)-astellatene, (–)-ent-quianulatene, (–)-variculatriene A, (–)-caprutriene, (+)-boleracene, (–)-aleurodiscalene A, (+)-caprutriene C, (–)-caprudine A, (+)-brarapadiene A, (–)-brarapadiene B, (–)-arathanadiene A, and (–)-arathanadiene B, and also characterized the corresponding biosynthetic enzymes (Figure 1). Interestingly, some of these compounds are originally found in microorganisms, suggesting the convergent evolution of biosynthetic genes.

Since sesterterpenes have more carbons than sesqui- and diterpenes do, their structure is more flexible and complicated, which makes the mechanistic investigation challenging. For example, two reasonable pathways can be proposed for (+)-astellatene formation (Figure 2). In path I, a 5/6/11 tricyclic skeleton is formed first, whereas path II involves the initial formation of a 5/12/5 tricyclic skeleton. Prior to this research, a labeling experiment was performed by Ye et al. in the biosynthesis of sesterfisherol, which is a similar sesterterpenoid. They could not experimentally distinguish between the 5/6/5 and 5/12/5 tricyclic skeleton formations because both paths are consistent with the labeling experiment results (Ye et al., 2015). Moreover, predicting a reasonable pathway based on presumable intermediates is difficult. For example, (+)-arathanatriene and (–)-caprutriene could be formed *via* path I, whereas (–)-variculatriene A and (–)-variculatriene B could be formed *via* path II, indicating that the experimental approach is not enough to answer this question. Accordingly, Huang et al. performed DFT calculations to assess the viability of the proposed biosynthetic pathways of (+)-astellatene and brarapadiene A (Huang et al., 2018).

Based on their computational results, both pathways are energetically viable; however, path I is inherently preferred since the highest energy barrier in path I is lower than that in path II by 6 kcal/mol. This study shows that DFT calculations can clarify the inherent reactivity of the carbocation intermediate in terpene biosynthesis, which could be useful for distinguishing the favored pathway. This combined experimental and computational approach is very useful, particularly when the reaction mechanism cannot be accessed solely using experiments.

Density Functional Theory Calculation Revealed That the Initial Conformation Affects the Regio- and Stereoselectivity

In terpene cyclization cascade, the computed inherent reactivity is consistent with the experimental result, indicating that enzymes do not precisely manipulate the reactions step by step. However, conventionally, various kinds of terpenoids are synthesized from common isoprenoid substrates. Therefore, some may wonder how terpene cyclases can control the regio- and stereoselectivity without interacting with the intermediate during the cyclization.

To investigate the origin of the terpenoid structural diversity, Sato et al. carried out a detailed comparative study of two structurally similar sesterterpenes: sesterfisherol and quianulatene, which were synthesized by NfSS and EvQS, respectively (Sato et al., 2018a). Based on their calculations, these sesterterpenoids are formed *via* the 5/12/5 tricyclic intermediate unlike for astellatene (Figure 2, path II). In addition, a careful comparison study clarified the conformational difference between the eight-membered rings in these two 5/12/5 intermediates. Here, we refer to the conformation of the quianulatene's 5/12/5 tricyclic intermediate (IM_Q) as conformation A, and that of sesterfisherol (IM_S) as conformation B.

They carried out “conformation swapping analysis” among these two sesterterpenoid intermediates (Figure 3). When IM_Q is in conformation B, 5/5/5-type triquinane formation proceeds, which was observed in sesterfisherol biosynthesis (Figure 3B). However, this triquinane formation reaction does not proceed inside EvQS, because conformation A is more stable than conformation B by 2.3 kcal/mol, and the activation energy of the 5/5/6-type condensation reaction is much lower. When IM_S is in conformation A, a 5/5/6-type condensation reaction proceeds, which is observed in quianulatene biosynthesis (Figure 3C). However, this reaction cannot proceed inside NfSS because the energy barrier is too high ($\Delta G^\ddagger = 26.6$ kcal/mol). This analysis suggests that the destination of the cyclization cascade is determined by the conformation of the intermediates. For each intermediate, the preferred conformation is different (A is preferred in quianulatene biosynthesis, and B in sesterfisherol biosynthesis), which could be attributed to the stereochemistry of the 5/11 condensation positions (*trans*-fused in IM_Q while *cis*-fused in IM_S).

Moreover, a further detailed comparison was carried out for the whole cyclization cascade, from which the initial conformation was found to be critical for the selectivity (Figure 4). The orientation of the methyl group of each double bond in the initial conformation determines the stereochemistry of the intermediate; accordingly, the conformation of the intermediate is automatically fixed, and the destination of the cascade is automatically set. In conclusion, terpene synthase appears to regulate the regio- and stereoselectivity by fixing the initial conformation.

Theozyme Calculation Identified the Key Residue for Sesterfisherol Biosynthesis

Generally, the main roles of terpene cyclases are (1) abstraction of pyrophosphate, (2) fixing the initial conformation (as was

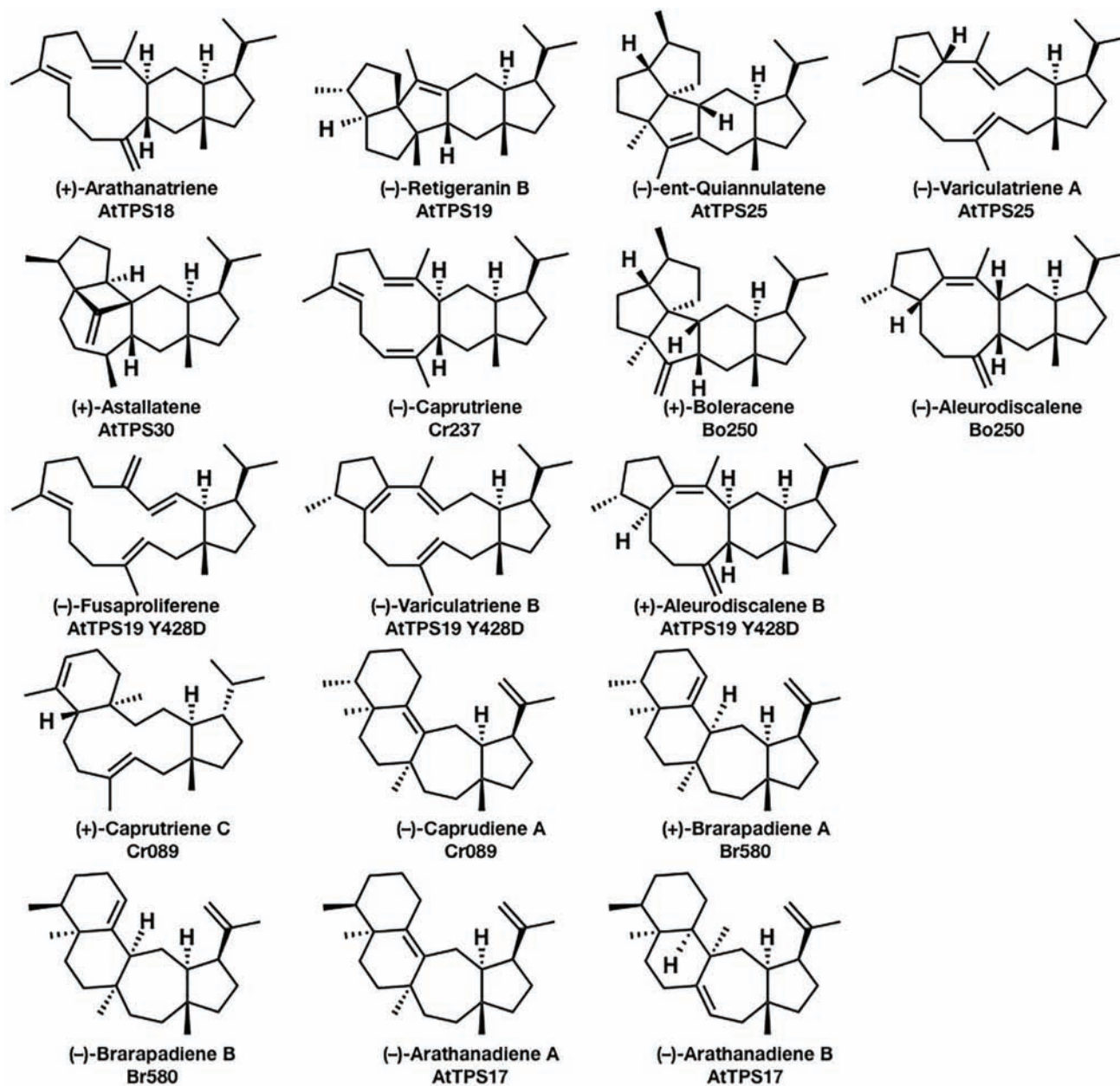


FIGURE 1 | Seventeen sesterterpenes were isolated from plants based on the genome mining approach. The compounds' names and their corresponding biosynthetic genes are shown below the structures. AtPTS is a terpene synthase from *Arabidopsis thaliana*; Cr is a terpene synthase from *Capsella rubella*; Bo is a terpene synthase from *Brassica oleracea*; and Br580 is a terpene synthase from *Brassica rapa*.

described above), (3) protecting the reactive intermediates from water, and (4) termination of the cyclization. The terpene-forming reaction is initiated by the elimination of the pyrophosphate group; afterward, the cyclization is driven by the inherent reactivity of the carbocation. However, the residue in the active site occasionally affects the carbocation intermediate in the cyclization cascade. Considering computational expenses, theozyme calculation is a great method for estimating the involvement of residues in the active site. Here, we show one example in which theozyme calculation and experimental validation worked successfully.

A detailed theoretical study of sesterfisherol biosynthesis was published by Sato et al. (Narita et al., 2017; Sato et al., 2018b). While they were exploring the cyclization mechanisms of sesterfisherol, a deep minima was found that required extraordinarily high energy, which was different from the activation energy in terpene biosynthesis. Subsequently, they examined the possibility of C–H π interactions (Hong and Tantillo, 2015), since terpene cyclase active sites are generally formed by aromatic and aliphatic residues. To estimate the C–H π interaction, a benzene ring was located around the

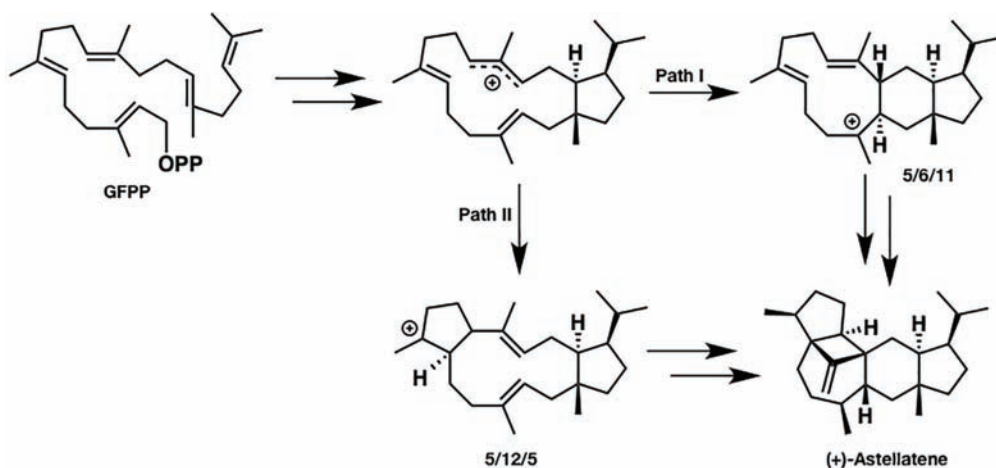


FIGURE 2 | Two reasonable biosynthetic pathways of (+)-astellatene. In path I, 5/6/11 tricyclic skeletons are formed. Conversely, a 5/12/5 tricyclic skeleton is formed in path II.

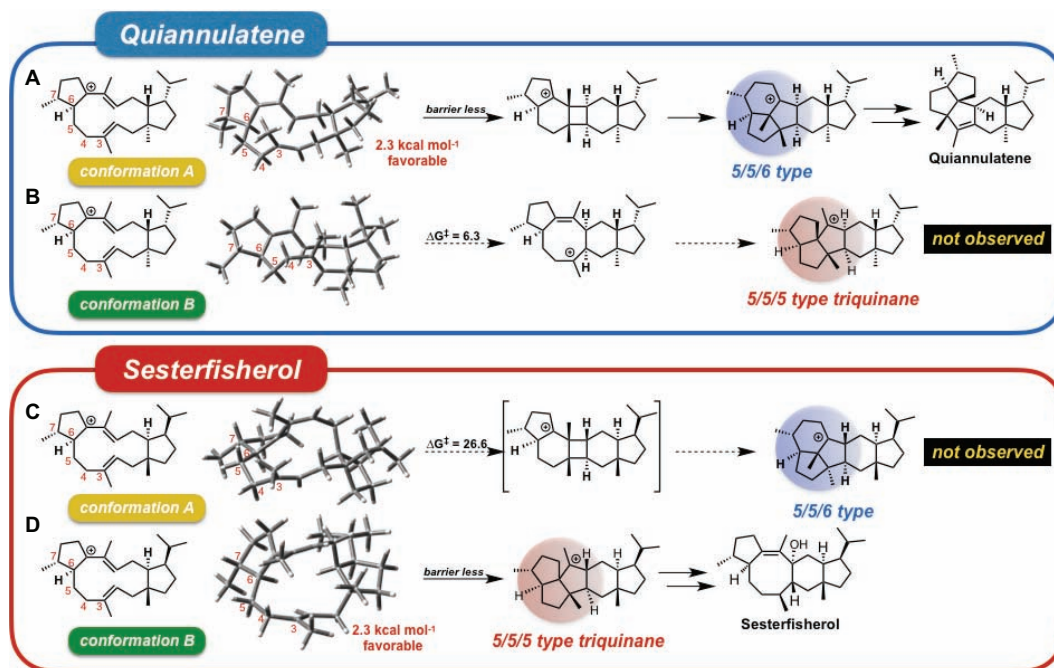


FIGURE 3 | Conformational swapping changed the regioselectivity. This figure represents the regioselectivity of the 5/11/5 tricyclic intermediates in both quiannulatene and sesterfisherol biosyntheses, in each conformation. **(A)** Quiannulatene biosynthesis pathway in conformation A, **(B)** Quiannulatene biosynthesis pathway in conformation B, **(C)** Sesterfisherol biosynthesis pathway in conformation A, **(D)** Sesterfisherol biosynthesis pathway in conformation B.

carbocation center of the intermediate, and the transition state structure search was subsequently carried out. Consequently, an alternative pathway was found, in which the C–H π interaction lowered the activation energy and also avoided the deep minima (Figure 5).

Notably, they also experimentally examined this C–H π interaction by site-directed mutagenesis. First, they constructed a homology model of NfSS and carried out the docking simulation. Afterward, several aromatic residue candidates

that could form the C–H π interaction were found in NfSS's active site (Figure 6). Due to their experiment, NfSS F191A produced new compounds instead of sesterfisherol (Figure 5). This result suggests that F191 is the key catalytic enzyme in sesterfisherol biosynthesis.

Interestingly, two of the products produced by NfSS F191A are thought to be derivatives of computationally predicted intermediates (Byproduct 1 and Byproduct 2), which could be deprotonated by the pyrophosphate adjacent to the active site.

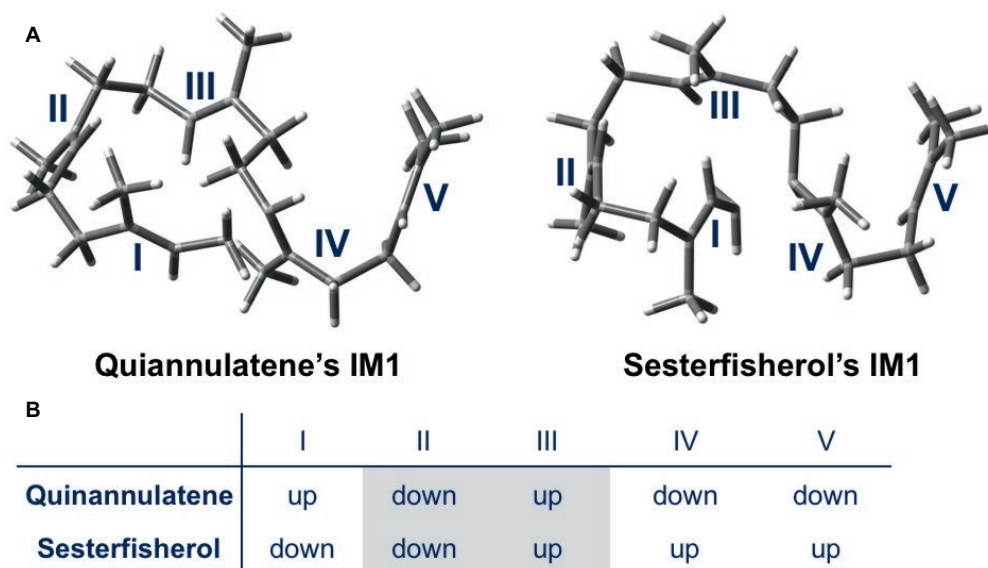


FIGURE 4 | Initial conformations of GFPP in quiannulatene and sesterfisherol biosyntheses. **(A)** 3D representations of IM1 in the quiannulatene biosynthesis pathway and the sesterfisherol biosynthesis pathway. **(B)** The orientation of methyl groups attached to the double bonds I, II, III, IV, and V.

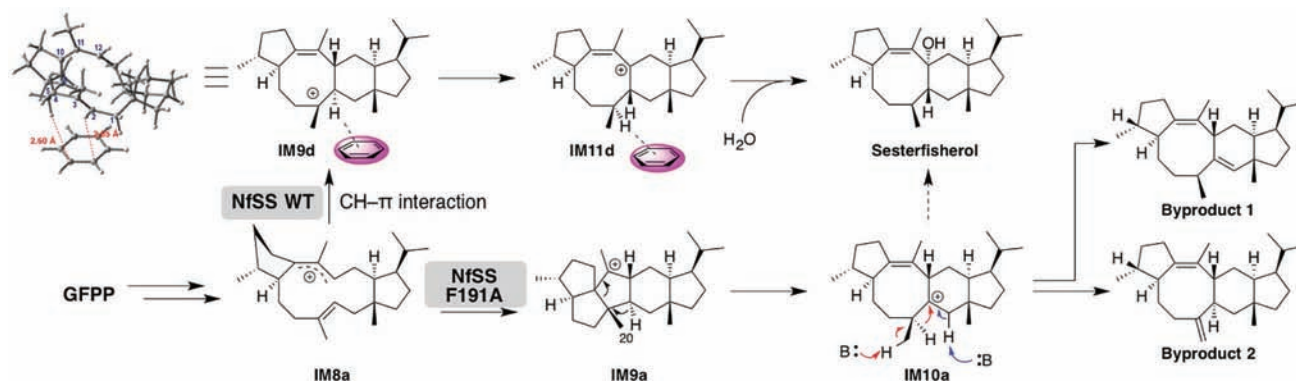


FIGURE 5 | Summary of the biosynthetic pathways catalyzed by NfSS WT and NfSS F191A.

As we have shown in this section, the combination of the theozyme calculation, homology modeling, docking simulation, and site-directed mutagenesis can be a powerful tool for finding the key catalytic residues. We think this approach can accelerate the mechanistic studies of biosynthetic enzymes in plant secondary metabolism.

High-Resolution All-Atom Model of Terpene Synthase

As was described above, terpene synthases play an important role in pre-organizing the initial conformation of the substrate. However, building molecular models of entire terpene-forming reactions within an active site remains a challenge. Major et al. mentioned the following in their review, “A crucial question in any study of terpene synthases is that of the correct binding mode. Indeed, crystal structures of terpene synthases often contain substrates bound in unreactive conformations, partly

due to the stickiness of the hydrocarbon moiety of the substrate and its lack of hydrogen bond potential. Thus, there is often great uncertainty regarding the correct binding mode when commencing multi-scale simulation projects of terpene cyclases” (Major et al., 2014). To tackle this problem, O’Brien et al. combined QM calculation and computational docking with Rosetta molecular modeling suite and reported the high-resolution all-atom models of epi-aristolochene synthase (TEAS) from *Tobacco* (O’Brien et al., 2016) and (+)-bornyl diphosphate synthase from *Salvia officinalis* (O’Brien et al., 2018).

Due to page limitations, we briefly explain their methodologies here (please refer to the original paper for more details). In the study, they initially carried out DFT calculations along with the generally accepted biosynthetic pathway of epi-aristolochene formation (Figure 7). All computed carbocation intermediates were subsequently subjected to conformational search using molecular mechanics force field (MMFF). These conformation

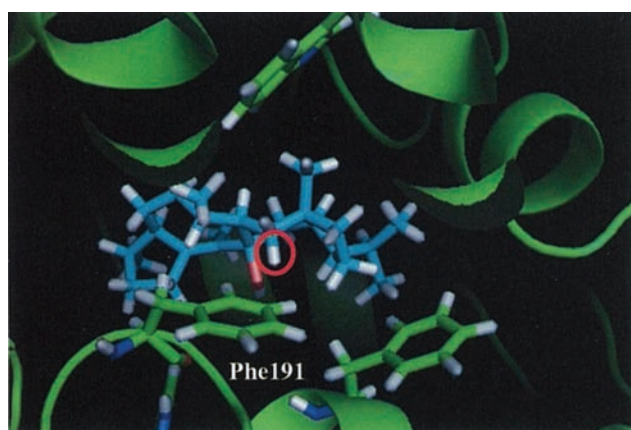


FIGURE 6 | A carbocation intermediate and NiFS complex obtained by the docking simulation.

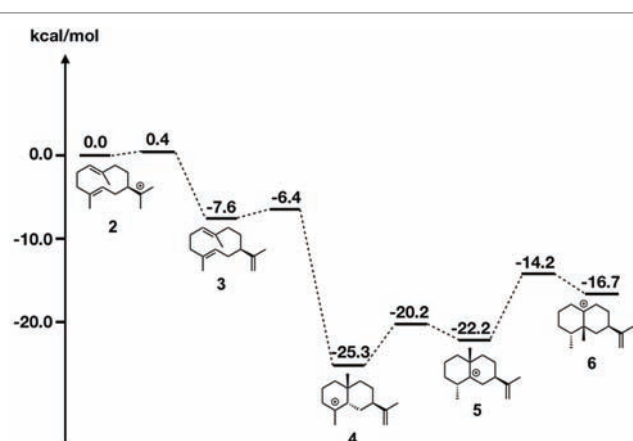


FIGURE 7 | Energy diagram of epi-aristolochene biosynthetic pathway. Relative energies to intermediate 2 are shown in kcal/mol.

libraries, containing over a hundred conformers for each carbocation intermediate, were subsequently optimized using DFT calculations, and low energy structures within 5 kcal/mol were used for the docking simulation. For the preparation of the protein structure, the X-ray crystal structure was minimized using a constrained FastRelax (Conway et al., 2014) procedure from the Rosetta modeling suite (Meiler and Baker, 2006; Richter et al., 2011). The diphosphate/magnesium complex extracted from the TEAS crystal structure was docked along with previously generated conformer libraries into the relaxed crystal structure using the chemically meaningful constraints. To ensure that the sampling was sufficient, 2,500 docking runs per catalytic orientation (motif) were performed. The resulting structures were combined and subsequently subjected to filtering based on three explicit constraints, involving (1) the departing diphosphate oxygen that results in the carbocation; (2) deprotonation of carbocation intermediate 2; and (3) protonation of carbocation intermediate 3.

As shown in **Figure 8**, motif 1 is abundant in all intermediates, suggesting that this orientation is the most reasonable docking

mode in TEAS. Furthermore, they carried out the RMSD calculation on the motif 1 structures in all intermediates and revealed the least movable docking mode during the cyclization reaction. Interestingly, only a few orientations are enriched in the early stage of this biosynthesis, whereas several motifs are enriched in the late stage of this enzyme reaction. In addition, the RMSD value was increased in the late stage of this cyclization reaction. These are consistent with the generally accepted concept that the substrate affinity decreases as the reaction proceeds. We think this method is applicable for future efforts to carry out the rational redesign of reaction specificity of this class of enzymes.

Quantum Mechanics/Molecular Mechanics Molecular Dynamics Calculation Revealed the Substrate Specificity of Geranyl Diphosphate Synthase

Although the inherent reactivities of the carbocation intermediates, computed by DFT calculations, are consistent with the experimental results, terpene cyclases occasionally affect the carbocation intermediates, as shown in Section “Theozyme Calculation Identified the Key Residue for Sesterfisherol Biosynthesis”. Therefore, the calculation including both the substrate and whole protein structure might be required to estimate the significance of the enzymatic support. For this purpose, the QM/MM method is widely used due to its computational cost. Here, we introduce a detailed study of geranyl diphosphate synthase (GPPS) from *Mentha piperita*, reported by Wu and Xu (Liu et al., 2014).

GPPS accepts isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DPP) as substrates and yields geranyl diphosphate (GPP), which is the first step of the chain elongation in isoprenoid biosynthesis. In the current study, they carried out MD simulations to reveal the mechanism of an “open-closed” conformation change of the catalytic pocket in the GPPS active site and identified a critical salt bridge between **Asp91** (in loop 1) and **Lys239** (in loop 2), which is responsible for opening or closing the catalytic pocket. In addition, the small subunit regulates the size and shape of the hydrophobic pocket to flexibly host substrates with different shapes and sizes (DPP/GPP/FPP, C₅/C₁₀/C₁₅). Furthermore, QM/MM MD simulations were carried out to explore the binding modes for the different substrates catalyzed by GPPS. GPPS is known to be a bifunctional enzyme and can catalyze GPP, GGPP, and a negligible amount of FPP formation. QM/MM MD simulation revealed that the distances and angles between two substrates are critical for the reaction. These parameters are similar when GPP or GGPP is produced, whereas the reverse is the case when FPP is produced. This study shows how QM/MM MD simulation is effective for clarifying the enzymatic effect toward the reaction in plant metabolism. Moreover, the key residues Asp91, Lys239, and Gln156, which could be good candidates for the site-directed mutagenesis, were found based on the computation.

As was described in Section “High-Resolution All-Atom Model of Terpene Synthase,” building the high-resolution all-atom model of terpene synthase has been challenging for a long time, which is one reason there are only a few examples of

intermediate	motif1	motif2	motif3	motif4	motif5	motif6	motif7	motif8	motif9
2	68.4			5.3			26.3		
3	72.8						23.4		3.7
5	83.0		0.4	16.2	0.4				
6	67.6	7.8	7.0	11.7	0.8	2.3	2.3		0.4
7	71.8	2.5	10.1	2.5	2.9		1.3		8.8

FIGURE 8 | Docking results using Rosetta. The structures of the intermediates are shown on the left. The darker the green in each box, the higher the percentage of low energy structures that are found in that catalytic orientation. If no low energy solutions were found for a particular intermediate, then no value was given. The number is the percentage of total low energy structures found for the catalytic motif when docking a particular intermediate.

QM/MM calculations of terpene synthase. We expect an increased number of reports on the all-atom modeling, using QM and Rosetta modeling suite, in the near future, which would promote QM/MM calculations. Moreover, we hope this kind of research will facilitate the mechanistic investigation and rational engineering of the biosynthetic pathways in plants.

ALKALOIDS

Unlike terpenoids, alkaloids have polar moieties that can form strong interactions with the enzyme. Therefore, more careful model construction is necessary, such as theozyme, MD simulation, or QM/MM. Accordingly, few theoretical studies on alkaloid biosynthesis have been reported to date. Here, we introduce four theoretical studies on the inherent reactivity in alkaloid biosynthesis.

Density Functional Theory Calculation Revealed a Favorable Mechanism for β -Carboline Formation Catalyzed by Strictosidine Synthase

Strictosidine is an important common intermediate that can be converted to many kinds of plant indole monoterpene alkaloids, such as ajmalicine, quinine, vinblastine, reserpine, camptothecin, and vincamine. The Pictet-Spengler condensation reaction of tryptamine with secologanin, catalyzed by strictosidine synthase, has been intensively studied, and several X-ray crystal structures have been reported. The Pictet-Spengler reaction consists essentially of two steps. First, an electron-rich aromatic amine attacks the aldehyde of secologanin to form an iminium intermediate. Second, an aryl amine attacks the electrophilic iminium to yield a positively charged intermediate that is then deprotonated to yield a β -carboline product. Although the ligand-bound crystal structure is available, the detailed mechanism of β -carboline formation remains unclear. Notably, carbon 2 and 3 of the indole moiety are nucleophilic; therefore, two possible mechanisms can be written (Figure 9). In path I, carbon 2 attacks the iminium moiety, after which the direct six-membered ring formation proceeds. In path II, carbon 2 attacks the iminium

moiety, thereby forming spiroindolenine, which is subsequently converted to β -carboline by a 1,2-alkyl shift. To clarify which pathway is more reasonable, Maresh et al. carried out DFT calculations (Maresh et al., 2008).

As shown in Figure 10, the formation of the six-membered ring is several orders of magnitude faster than spiroindolenine formation, which is consistent with the empirical rule: that 6-endo-ring closures are favored over 5-endo-trig cyclization. However, the formation of the spiroindolenine intermediate has been observed by isotopic scrambling. In addition, spiroindolenine formation requires only 8 kcal/mol, and it appears to occur under ambient conditions. Therefore, if the deprotonation is slow enough, spiroindolenine can be formed during the course of the reaction. Moreover, this calculation also suggests that the 1,2-alkyl shift that connects iminium spiroindolenine and the six-membered ring intermediate requires significantly high energy, which does not contribute to the mechanism.

Even though we obtained X-ray crystal structures, revealing the detailed reaction mechanisms is often challenging. This study indicates that the combined strategy of computations and kinetic isotope experiments is quite effective in distinguishing two possible pathways. We hope this approach will be widely utilized in the study of plant secondary metabolism.

Camptothecin E Ring Opening Reaction

Camptothecin (CPT) is a plant alkaloid that was originally isolated from the Chinese tree, *Camptotheca acuminata*, in 1966 (Wall et al., 1966). It features a planar pentacyclic ring structure with a pyrrolo [3-4- β] quinolone moiety (rings A, B, and C), a conjugated pyridine group (ring D), and one α -hydroxy lactone ring (ring E); consequently, it has received interest from scientists due to its remarkable anticancer activity in preliminary clinical trials. Although two CPT analogs, topotecan and irinotecan, have been approved and are currently used in cancer chemotherapy, the pharmacological investigation into CPT had been suspended on one occasion due to the poor solubility of CPT in water and most organic solvents (Gupta et al., 1995; Wu and Liu, 1997; Wall, 1998).

The stabilities of CPT and its derivatives in solution were reported to be highly pH-dependent by several experimental

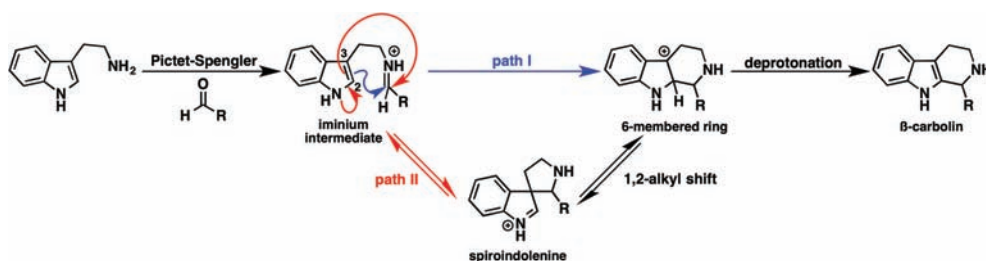


FIGURE 9 | Possible pathways for β -carboline formation.

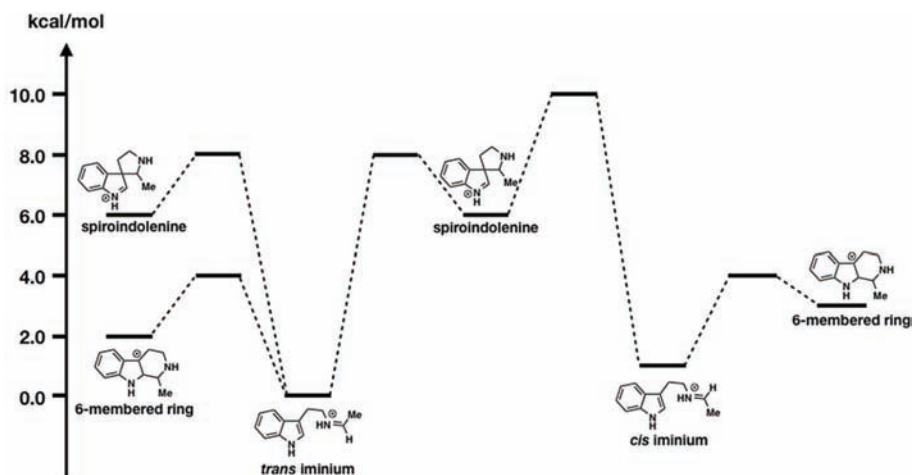


FIGURE 10 | The energy diagrams of the six-membered ring formation from spiroindolenine. Relative free energies to trans iminium are shown in kcal/mol.

analyses. In addition, CPT and its derivatives tend to aggregate in solution, particularly in dimer formation. One of the key factors is the E-ring-opening reaction that can proceed at neutral pH. Zou et al. carried out DFT calculations to obtain the theoretical basis for this E ring opening reaction (Zou et al., 2013). In their study, they revealed (1) that the E-ring opening reaction can proceed under the physiological pH ($\Delta G^\ddagger = 12.94$ in aqueous), (2) the solvation effect, and (3) the substitution effect. This study is a good example of how computational chemistry is an effective tool for revealing the degradation process of plant alkaloids and for examining their instability.

Biosynthetic Dipolar Cycloaddition: Daphniphyllum Alkaloids, Flueggeine A, and Virosaines

Cycloaddition is one of the most important reactions for skeletal formation in both total synthesis and biosynthesis of natural products. There have been many theoretical studies on Diels-Alderase which provide insightful mechanistic approaches that cannot be easily addressed by experimental approaches. These reports also show that the inherent reactivity of a molecule is relevant to many cyclization reactions promoted by enzymes. Here, we show two theoretical studies of plant alkaloids from *Daphniphyllum* (Tantillo, 2016) and *Flueggea* (Painter et al., 2013).

The development of cascade polycyclizations by Heathcock and co-workers, to construct *Daphniphyllum* alkaloids, is a milestone in biomimetic total synthesis, in which bicyclic intermediate A is thought to be converted into tetracyclic intermediate D via intermediates B and C (Heathcock et al., 1992; Heathcock, 1996). Based on their calculations, A is converted to B with an activation energy of only 7.1 kcal/mol via concerted [4 + 2] cycloaddition (aza-Diels-Alder reaction). This result represents a potential biological Diels-Alder reaction for which enzymatic barrier lowering would not be required. It also demonstrates that enzymatic preorganization of the substrate is not required for a successful reaction, presumably because most conformers are unreactive. Interestingly, B is directly converted to D via ene reaction, in which C is not a minima but a transient structure (Figure 11).

While there have been many studies on [4 + 2] cycloaddition reactions, other types of cycloadditions received less attention. Flueggeine A and virosaine alkaloids, isolated from *Flueggea virosa*, are thought to be synthesized from a common precursor that could be derived from norsecurinine or tyrosine (Figure 12). Interestingly, this proposed biosynthetic pathway involves a nitron-alkene (3 + 2) cycloaddition reaction.

In the referenced study, they examined all possible eight stereoisomers for flueggeine A formation. The predicted energy barrier for the cycloaddition in water via this transition state

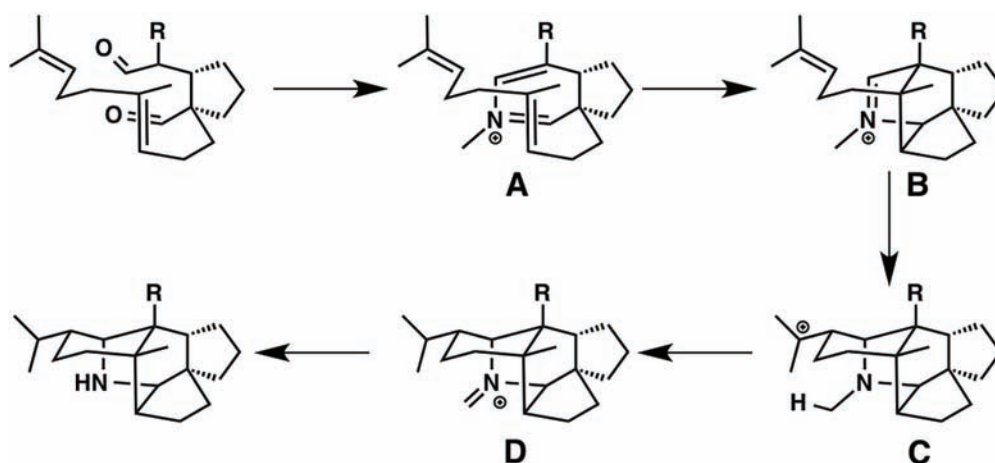


FIGURE 11 | Biosynthetic pathway of *Daphniphyllum* alkaloids.

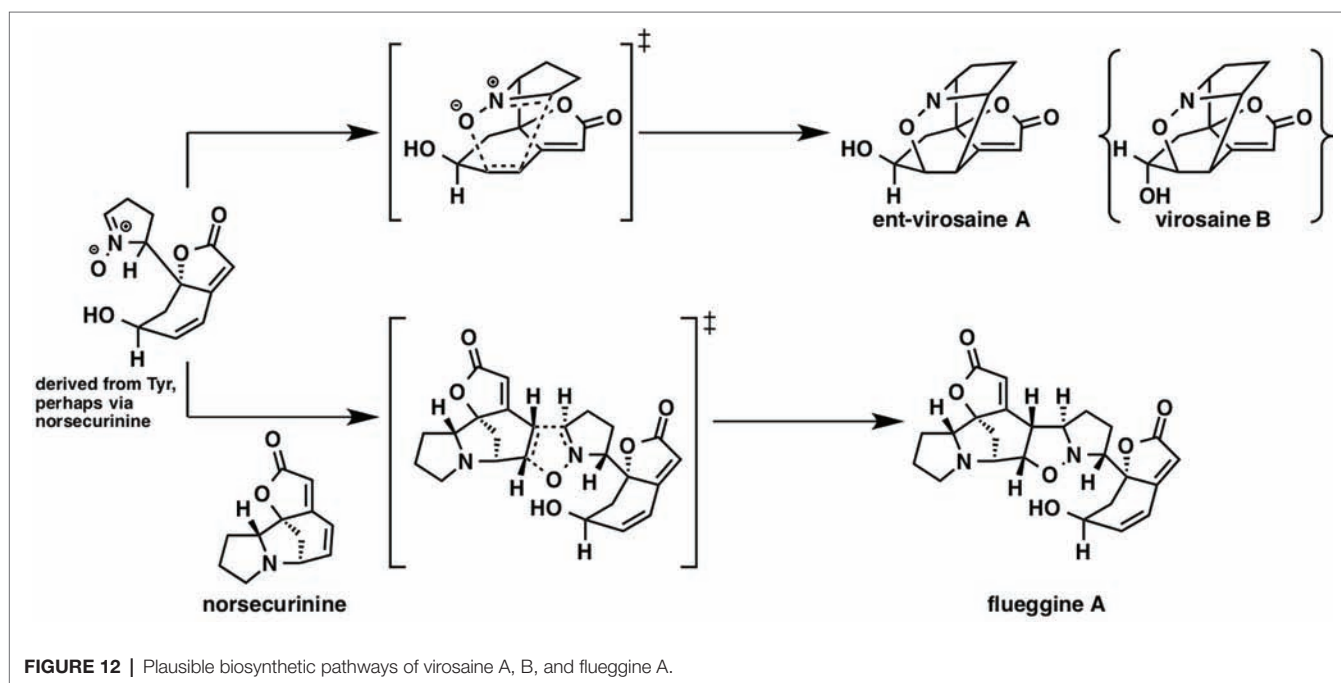


FIGURE 12 | Plausible biosynthetic pathways of virosaine A, B, and flueggine A.

structure is 10.3 kcal/mol, which is lowest among all the transition state structures for this reaction, implying that selectivity control by the enzyme is not required. Based on the distortion interaction model, this transition state has both considerably small distortion energy and considerably large interaction energy. They also examined virosaine formation, and the activation energy was only ~1 kcal/mol, indicating that enzymatic assistance is unnecessary.

It is always difficult to ascertain if the Diels-Alder reaction involved in the proposed biosynthetic pathway is enzymatic or not. However, computational chemistry provides us with the energy landscape of the reaction, which can assist in narrowing down the candidate biosynthetic genes in plant science.

Elucidation of Δ^1 -Piperidine Dimer Formation in Quinolizidine Alkaloid Biosynthesis

Unlike for fungi and bacteria, revealing whole biosynthetic genes relevant to a plant secondary metabolite is still quite challenging. Quinolizidine alkaloids (QAs), a subclass of Lys-derived plant alkaloids widely distributed in *Leguminosae*, have been intensively studied for over half a century. However, most of their biosynthetic genes, enzymes, and intermediates remain unknown. Particularly, in the quinolizidine skeletal formation process, which is thought to be common in all QA biosynthesis, only two enzymes, L-lysine decarboxylase (LDC) (Bunsupa et al., 2012) and copper amine oxidase (CAO) (Yang et al., 2017), have been identified to date (Figure 13A).

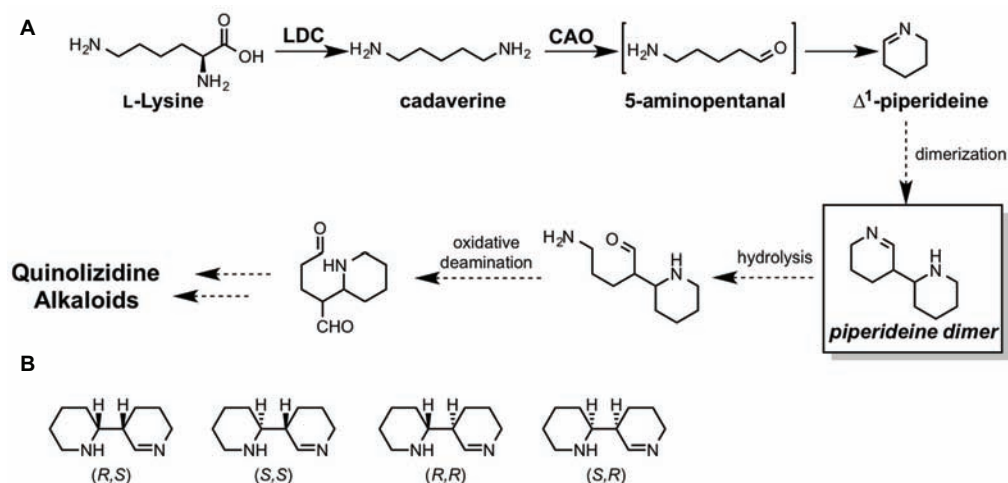


FIGURE 13 | (A) proposed biosynthetic pathways of quinolizidine alkaloids. LDC stands for lysine decarboxylase, and CAO stands for copper amine oxidase. **(B)** Four possible stereoisomers of piperidine dimers.

Therefore, there is no valid information on how many genes, enzymes, and reactions are required for production of QAs after Δ^1 -piperidine formation. In a plausible biosynthetic pathway of QAs, 5-aminopentanal is spontaneously converted to Δ^1 -piperidine (**Figure 13A**). The dimerization of Δ^1 -piperidine has been proposed; however, it is unclear whether or not it requires enzymatic assistance. Therefore, Sato et al. carried out theoretical investigations to uncover the biosynthetic mechanism of Δ^1 -piperidine dimerization (Sato et al., 2018c). There are four possible stereoisomers for the piperidine dimer, i.e., (R,R), (S,S), (R,S), and (S,R) (**Figure 13B**). In the mentioned study, they constructed a model that has two molecules of Δ^1 -piperidine and two to four molecules of water for each isomer.

The results of the DFT calculations indicate that piperidine dimerization spontaneously proceeds under neutral conditions and yields only (R,R) or (S,S) dimers. (R,S) and (S,R) dimer formations require considerably higher energy barriers; therefore, they cannot be formed under neutral conditions. Based on the previous literature, tetrahydroanabasin, which could be formed from the (R,R) piperidine dimer through isomerization, was isolated from a QA producing plant. Therefore, (R,R) piperidine formation could be favored in a QA producing plant. This stereochemistry is also consistent with the other QA derivatives. For example, (–)-lupinine and (+)-epilupinine could be synthesized from (R,R)-piperidine and (R,S)-piperidine dimers, respectively. However, we cannot rule out the possibility that the enzyme assists this dimerization reaction since the (S,S)-piperidine dimer has not yet been isolated. Moreover, nature occasionally uses an enzyme for the reaction that can spontaneously proceed under aqueous conditions (Chen et al., 2018).

In the study of QA biosynthesis, differential expression analysis was performed to search for LDC genes, which also provided dozens of other candidate genes. The combination of computational chemistry and expression-based approach could be a significantly powerful tool for narrowing down the candidate genes.

OTHER SECONDARY METABOLITES

Unlike the case with other natural product groups, there are only a few mechanistic investigations on flavonoid and lignin biosynthesis, although there are many studies on spectrum prediction regarding flavonoid. Here, we introduce one study about anthocyanin biosynthesis and three studies about lignin biosynthesis.

A Long-Standing Issue in Anthocyanin Biosynthesis Was Solved Using Density Functional Theory Calculations

The biosynthetic pathways of flavonoids have been intensively studied, and related genes, enzymes, and intermediates have been characterized. In the late stage of the biosynthesis of anthocyanin, dihydroflavonoid is converted to anthocyanidin, which is catalyzed by two enzymes, dihydroflavonol reductase (DFR) (Gong et al., 1997) and anthocyanidin synthase (ANS) (**Figure 14**, Route A; Saito et al., 1999; Nakajima et al., 2001; Turnbull et al., 2004). Interestingly, reduction proceeds just after the oxidation in this conversion, which appears to be a detour pathway and energy consuming, because the oxidation state does not change after these two reactions. An alternative non-enzymatic pathway, that is a simple spontaneous tautomerization, was proposed (**Figure 14**, route B). However, this spontaneous pathway does not proceed inside the plant because the suppression of genes leads to a decrease in the amount of anthocyanidin. Therefore, the question about this biosynthetic pathway is why this tautomerization pathway is not feasible. To answer this question, Sato et al. carried out DFT calculations (Sato et al., 2018d).

Based on their calculations, the first tautomerization *via* the non-enzymatic pathway requires ca. 30 kcal/mol, which is too high to be achieved under ambient conditions. This is due to the instability of the transition state structure that has a highly electron rich en-diol moiety adjacent to the electron-rich aromatic ring. Interestingly, dihydroflavonol without hydroxyl

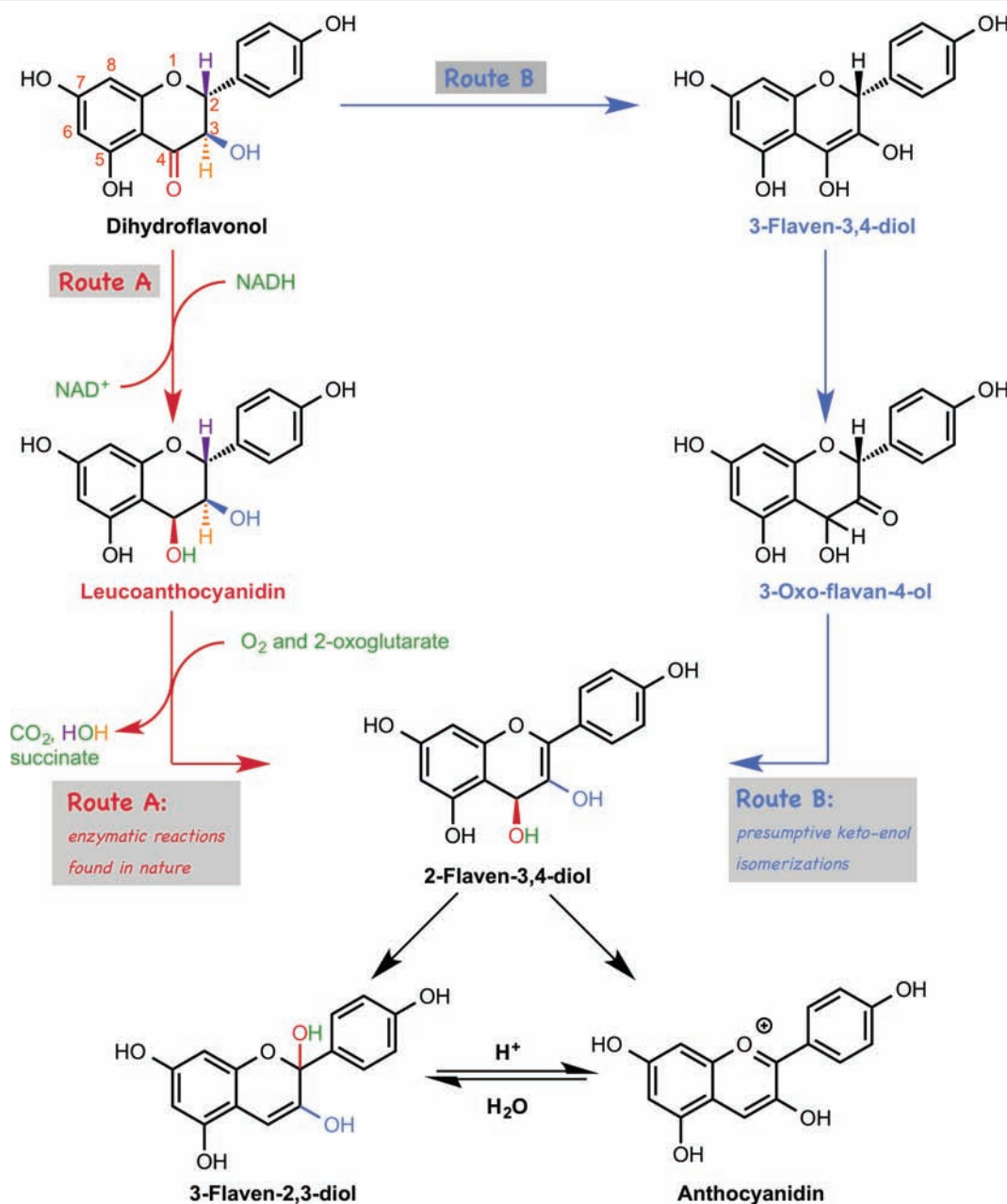


FIGURE 14 | Two possible pathways for anthocyanidin biosynthesis. In route A, the conversion of dihydroflavonol to anthocyanidin is catalyzed by two enzymes, such as dihydroflavonol reductase (DFR) and anthocyanidin synthase (ANS). In route B, dihydroflavonol is converted to anthocyanidin by a three-step tautomerization reaction.

groups at the 5' and 7' positions require lower activation energies than the original one does, indicating that the hydroxyl group is the natural controller of the yield of anthocyanidin. Moreover, the first tautomerization requires less activation energy under the acidic conditions, suggesting that this tautomerization requires enzymatic support. They searched for enzymes that could catalyze this type of tautomerization using the KEGG database, but no enzyme was found.

Furthermore, they investigated the later part of the anthocyanin biosynthetic pathway. However, interestingly, their DFT calculation

suggests that 2-flaven-3,4-diol is directly converted to anthocyanidin under acidic conditions, which was originally thought to be converted *via* 3-flaven-2,3-diol formation (Figure 14). This new finding is consistent with the experimental result in the previous literature, in which hydrochloride was used to terminate the ANS reaction, and 3-flaven-2,3-diol was not detected.

As we described here, computational chemistry can simulate a reaction that cannot actually occur and can also easily delete the functional groups to assess their effect on the reaction, which is quite difficult in organic synthesis. This study emphasizes

the usefulness of the computational approach to test the hypothetical pathways.

Lignin

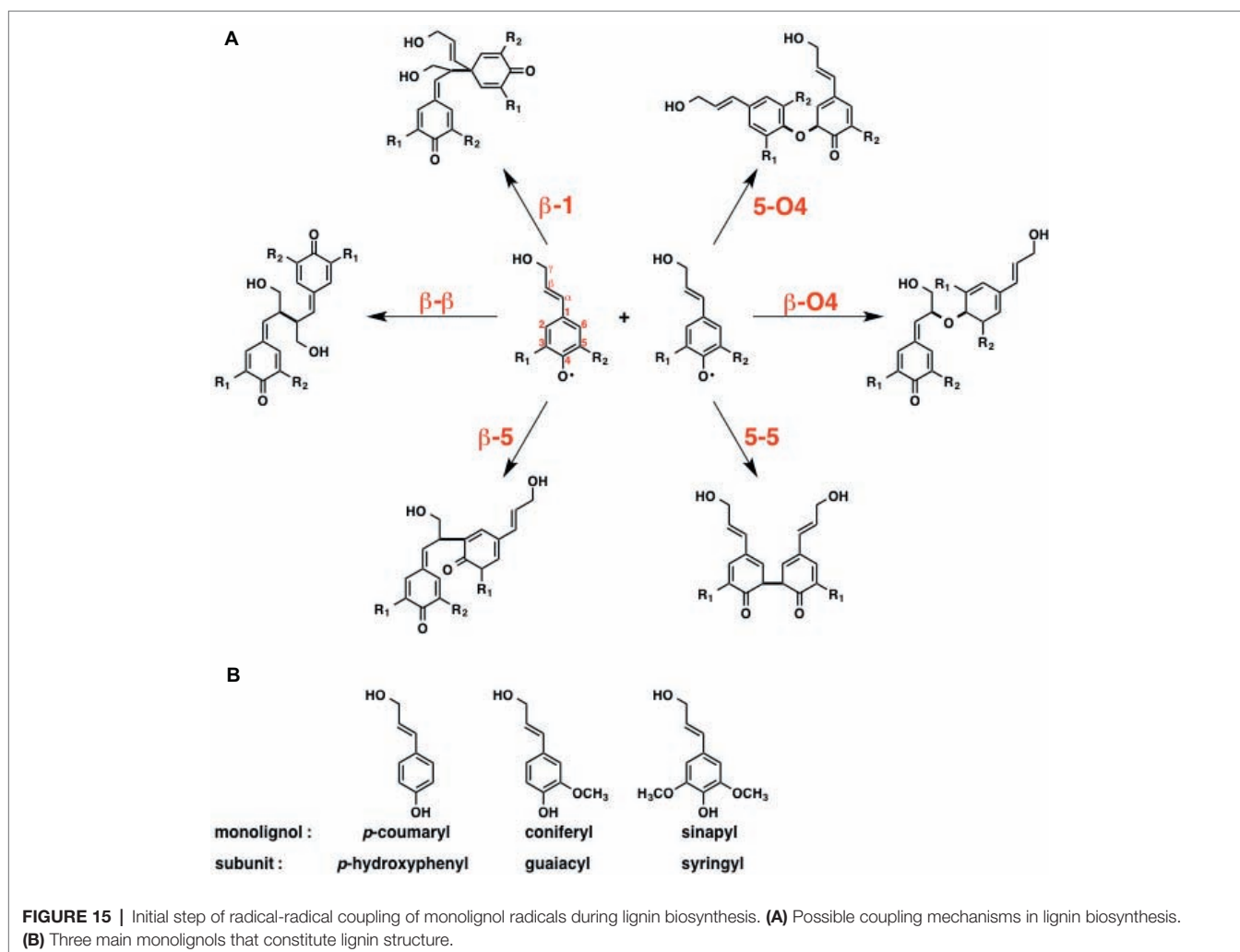
Unlike the other types of natural products that are mentioned above, only a few examples were reported for the mechanistic investigation of lignin biosynthesis. The theoretical investigation of lignin biosynthesis is challenging for some reasons, i.e., radical reaction, branched polymer structure, various kinds of coupling products, etc. To reveal the mechanism of lignin biosynthesis, a consecutive theoretical study has been reported by Sangha et al. (2012, 2014, 2016).

The initial step of lignin biosynthesis is initiated by the radical coupling reaction of monolignols. Besides the reactive oxygen, the monolignol derived radicals are reactive at the C1, C3, C5, and β positions due to the delocalization *via* the conjugated π system. Conventionally, lignin polymerization, catalyzed by peroxidases and laccases, takes place in three steps: (1) monolignol binding to the enzyme active site, (2) H_2O_2 -mediated oxidation at the active site to form radicals, and (3) radical coupling reaction to form lignin polymers.

To reveal the inherent reactivity of monolignol, Sangha et al. initially carried out DFT calculation for the six possible radical coupling products, as shown in **Figure 15A**. The result indicates that the formation of β -O4, β - β , and β -5 type monolignol radical couplings is enthalpically favored over those of others, which is consistent with the experimental data suggesting that these three bonds are most common in natural lignin.

Next, they focused on horseradish peroxidase C (HRPC), which triggers the radical polymerization cascade in lignin biosynthesis. Lignin's subunits can be classified into three groups; guaiacyl (G), syringyl (S), and *p*-hydroxyphenyl (H) (**Figure 15B**). The ratio of the subunits is relevant to the efficiency of the deconstruction of biomass. Therefore, they examined the binding affinity of three monolignols, including *p*-coumaryl, coniferyl, and sinapyl, toward the horseradish peroxidase C (HRPC). To answer this question, they carried out MD simulations. The results indicated that the binding affinity of the monolignols toward HRPC decreases in the order of *p*-coumaryl > sinapyl > coniferyl alcohol.

Since lignin biosynthesis is a combination of simple radical coupling reactions, many possible reaction pathways can be proposed. As we have shown in this section, computational chemistry can reveal the inherent reactivity and can provide



a rational explanation for substrate specificity, which is helpful for narrowing down the candidate structures. We believe that computational chemistry can eliminate the hindrances to precise mechanistic investigations of radical coupling reactions in lignin biosynthesis.

SUMMARY AND PERSPECTIVES

As we have shown in this review, computational chemistry can be a powerful tool for revealing the biosynthesis of secondary metabolites in plants. Unlike cheminformatics, computational chemistry provides not only reasonable reaction pathways and energy barriers but also many new insights, as described above. For example, ring cyclization order in astellatene biosynthesis, key catalytic residues of biosynthetic enzymes, the controlling mechanism of regio- and stereoselectivity in quianulatene biosynthesis, docking mode in the active site of terpene synthase, detailed reaction mechanisms of piperidine dimerization, etc., which cannot be achieved solely by traditional experimental methods.

Although computational approaches using QM calculation, MD simulation, and QM/MM are well established in terpene-forming reactions and cycloadditions, only a few preliminary studies on other types of plant secondary metabolism have been reported until now. Many studies on other types of natural products are required to sophisticate this powerful computational approach. Particularly, more reports on the oxidation reactions, mainly catalyzed by P450, FMO, or iron-dependent enzymes (Nakashima et al., 2018), are desired because oxidation is one of the key reactions facilitating the structural diversity and complexity of plant secondary metabolites.

As was mentioned in Section “Introduction,” the next goal of these applicable studies using computational chemistry in natural product biosynthesis is to engineer or design novel biosynthetic pathways and enzymes to obtain desired products. One approach to achieve this objective is swapping the biosynthetic enzymes (genes) with other genes that can accept the biosynthetic intermediate, which could produce novel natural products. Another approach could be to design novel enzymes by computational chemistry (Kiss et al., 2013). The methodology of *de novo* enzyme design was reported by a decade ago, although it has not yet been applied to plant

secondary metabolism engineering. They published three remarkable examples of artificially designed enzymes: retroaldolase (Jiang et al., 2008), Kemp elimination enzyme (Röthlisberger et al., 2008), and Diels-Alderase (Siegel et al., 2010). In their method, they initially designed the catalytic cycle. Subsequently, they carried out DFT (theozyme) calculations to obtain the transition state structures. The next step involved searching the template protein, which has enough space in its cavity to accommodate the transition state structure. Finally, the enzyme was designed using Rosetta, which is also used in the terpene docking simulation, as described above. They successfully designed a desired artificial enzyme that was nonexistent. Moreover, they successfully improved the catalytic ability of those enzymes using directed evolution for several generations. We believe this kind of approach is very useful for the efficient production of novel bioactive plant secondary metabolites. We think terpene synthase is a good candidate for designing the artificial enzyme, since (1) the computational approach is well developed, (2) many transition-state structures are already available, (3) substrate is commercially available, and (4) most of the terpenes have unique biological activities. We hope that more biosynthetic studies utilize a computational chemistry-based approach for mechanistic investigations.

AUTHOR CONTRIBUTIONS

HS wrote entire manuscript under the supervision of KS and MY.

FUNDING

This work is supported by the JSPS Grant-in-Aid for Scientific Research on Innovative Areas JP16H06454 (MY).

ACKNOWLEDGMENTS

We gratefully acknowledge the Strategic Priority Research Promotion Program of the Chiba University. We would also like to thank *Editage* (www.editage.jp) for English language editing.

REFERENCES

- Afendi, F. M., Okada, T., Yamazaki, M., Hirai-Morita, A., Nakamura, Y., Nakamura, K., et al. (2011). KNApSACk family databases: integrated metabolite-plant species databases for multifaceted plant research. *Plant Cell Physiol.* 53:e1. doi: 10.1093/pcp/pcr165
- Bunsupa, S., Katayama, K., Ikeura, E., Oikawa, A., Toyooka, K., Saito, K., et al. (2012). Lysine decarboxylase catalyzes the first step of quinolizidine alkaloid biosynthesis and coevolved with alkaloid production in leguminosae. *Plant Cell* 24, 1202–1216. doi: 10.1105/tpc.112.095885
- Chen, X., Hagel, J. M., Chang, L., Tucker, J. E., Shiigi, S. A., Yelapaala, Y., et al. (2018). A pathogenesis-related 10 protein catalyzes the final step in thebaine biosynthesis. *Nat. Chem. Biol.* 14, 1–12. doi: 10.1038/s41589-018-0059-7
- Christianson, D. W. (2017). Structural and chemical biology of terpenoid cyclases. *Chem. Rev.* 117, 11570–11648. doi: 10.1021/acs.chemrev.7b00287
- Conway, P., Tyka, M. D., DiMaio, F., Konerding, D. E., and Baker, D. (2014). Relaxation of backbone bond geometry improves protein energy landscape modeling. *Protein Sci.* 23, 47–55. doi: 10.1002/pro.2389
- Dickschat, J. S. (2016). Bacterial terpene cyclases. *Nat. Prod. Rep.* 33, 87–110. doi: 10.1039/C5NP00102A
- Fukui, K. (1981). The path of chemical reactions – the IRC approach. *Acc. Chem. Res.* 14, 363–368. doi: 10.1021/ar00072a001
- Galanie, S., Thodey, K., Trenchard, I. J., Interrante, M. F., and Smolke, C. D. (2015). Complete biosynthesis of opioids in yeast. *Science* 349, 1095–1100. doi: 10.1126/science.aac9373
- Gong, Z., Yamazaki, M., Sugiyama, M., Tanaka, Y., and Saito, K. (1997). Cloning and molecular analysis of structural genes involved in anthocyanin biosynthesis

- and expressed in a forma-specific manner in *Perilla frutescens*. *Plant Mol. Biol.* 35, 915–927. doi: 10.1023/A:1005959203396
- Gonzalez, C., and Schlegel, H. B. (1989). An improved algorithm for reaction path following. *J. Chem. Phys.* 90, 2154–2161. doi: 10.1063/1.456010
- Gupta, M., Fujimori, A., and Pommier, Y. (1995). Eukaryotic DNA topoisomerases I. *Biochim. Biophys. Acta* 1262, 1–14. doi: 10.1016/0167-4781(95)00029-G
- Heathcock, C. H. (1996). Nature knows best: an amazing reaction cascade is uncovered by design and discovery. *Proc. Natl. Acad. Sci. USA* 93, 14323–14327. doi: 10.1073/pnas.93.25.14323
- Heathcock, C. H., Hansen, M. M., Ruggeri, R. B., and Kath, J. C. (1992). Daphniphyllum alkaloids. 11. Biomimetic total synthesis of methyl homosecodaphniphyllate. Development of the tetracyclization reaction. *J. Org. Chem.* 57, 2544–2553. doi: 10.1021/jo00035a008
- Hong, Y. J., and Tantillo, D. J. (2015). Tension between internal and external modes of stabilization in carbocations relevant to terpene biosynthesis: modulating minima depth via C–H... π interactions. *Org. Lett.* 17, 5388–5391. doi: 10.1021/acs.orglett.5b02740
- Huang, A. C., Hong, Y. J., Bond, A. D., Tantillo, D. J., and Osbourn, A. (2018). Diverged plant terpene synthases reroute the carbocation cyclization path towards the formation of unprecedented 6/11/5 and 6/6/7/5 sesterterpene scaffolds. *Angew. Chem. Int. Ed.* 57, 1291–1295. doi: 10.1002/anie.201711444
- Huang, A. C., Kautsar, S. A., Hong, Y. J., Medema, M. H., Bond, A. D., Tantillo, D. J., et al. (2017). Unearthing a sesterterpene biosynthetic repertoire in the Brassicaceae through genome mining reveals convergent evolution. *Proc. Natl. Acad. Sci. USA* 114, E6005–E6014. doi: 10.1073/pnas.1705567114
- Ishida, K., Morokuma, K., and Komornicki, A. (1977). The intrinsic reaction coordinate. An ab initio calculation for $\text{HNC} \rightarrow \text{HCN}$ and $\text{H}^+ + \text{CH}_4 \rightarrow \text{CH}_3 + \text{H}$. *J. Chem. Phys.* 66, 2153–2156. doi: 10.1063/1.434152
- Jensen, F. (2017). *Introduction to computational chemistry*. 3rd edn. (New York City, United States: Wiley).
- Jiang, L., Althoff, E. A., Clemente, F. R., Doyle, L., Röthlisberger, D., Zanghellini, A., et al. (2008). De novo computational design of retro-aldol enzymes. *Science* 319, 1387–1391. doi: 10.1126/science.1152692
- Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45, D353–D361. doi: 10.1093/nar/gkw1092
- Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27
- Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K., and Tanabe, M. (2019). New approach for understanding genome variations in KEGG. *Nucleic Acids Res.* 47, D590–D595. doi: 10.1093/nar/gky962
- Kiss, G., Çelebi-Ölçüm, N., Moretti, R., Baker, D., and Houk, K. N. (2013). Computational enzyme design. *Angew. Chem. Int. Ed.* 52, 5700–5725. doi: 10.1002/anie.201204077
- Liu, Z., Zhou, J., Wu, R., and Xu, J. (2014). Mechanism of assembling isoprenoid building blocks 1. Elucidation of the structural motifs for substrate binding in geranyl pyrophosphate synthase. *J. Chem. Theory Comput.* 10, 5057–5067. doi: 10.1021/ct500607n
- Luo, X., Reiter, M. A., d’Espaux, L., Wong, J., Denby, C. M., Lechner, A., et al. (2019). Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nature* 567, 1–18. doi: 10.1038/s41586-019-0978-9
- Major, D. T., Freud, Y., and Weitman, M. (2014). Catalytic control in terpenoid cyclases: multiscale modeling of thermodynamic, kinetic, and dynamic effects. *Curr. Opin. Chem. Biol.* 21, 25–33. doi: 10.1016/j.cbpa.2014.03.010
- Mareš, J. J., Giddings, L.-A., Friedrich, A., Loris, E. A., Panjikar, S., Trout, B. L., et al. (2008). Strictosidine synthase: mechanism of a Pictet-Spengler catalyzing enzyme. *J. Am. Chem. Soc.* 130, 710–723. doi: 10.1021/ja077190z
- Matsuda, S. P. T., Wilson, W. K., and Xiong, Q. (2006). Mechanistic insights into triterpene synthesis from quantum mechanical calculations. Detection of systematic errors in B3LYP cyclization energies. *Org. Biomol. Chem.* 4, 530–514. doi: 10.1039/b513599k
- Meiler, J., and Baker, D. (2006). ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility. *Proteins* 65, 538–548. doi: 10.1002/prot.21086
- Nakajima, J. I., Tanaka, Y., Yamazaki, M., and Saito, K. (2001). Reaction mechanism from leucoanthocyanidin to anthocyanidin 3-glucoside, a key reaction for coloring in anthocyanin biosynthesis. *J. Biol. Chem.* 276, 25797–25803. doi: 10.1074/jbc.M100744200
- Nakashima, Y., Mitsuhashi, T., Matsuda, Y., Senda, M., Sato, H., Yamazaki, M., et al. (2018). Structural and computational bases for dramatic skeletal rearrangement in anditomin biosynthesis. *J. Am. Chem. Soc.* 140, 9743–9750. doi: 10.1021/jacs.8b06084
- Narita, K., Sato, H., Minami, A., Kudo, K., Gao, L., Liu, C., et al. (2017). Focused genome mining of structurally related sesterterpenes: enzymatic formation of enantiomeric and diastereomeric products. *Org. Lett.* 19, 6696–6699. doi: 10.1021/acs.orglett.7b03418
- O’Brien, T. E., Bertolani, S. J., Tantillo, D. J., and Siegel, J. B. (2016). Mechanistically informed predictions of binding modes for carbocation intermediates of a sesquiterpene synthase reaction. *Chem. Sci.* 7, 4009–4015. doi: 10.1039/C6SC00635C
- O’Brien, T. E., Bertolani, S. J., Zhang, Y., Siegel, J. B., and Tantillo, D. J. (2018). Predicting productive binding modes for substrates and carbocation intermediates in terpene synthases-bornyl diphosphate synthase as a representative case. *ACS Catal.* 8, 3322–3330. doi: 10.1021/acscatal.8b00342
- Paddon, C. J., Westfall, P. J., Pitera, D. J., Benjamin, K., Fisher, K., McPhee, D., et al. (2013). High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* 496, 528–532. doi: 10.1038/nature12051
- Page, M., Doubleday, C., and McIver, J. W. (1990). Following steepest descent reaction paths. The use of higher energy derivatives with ab initio electronic structure methods. *J. Chem. Phys.* 93, 5634–5642. doi: 10.1063/1.459634
- Painter, P. P., Pemberton, R. P., Wong, B. M., Ho, K. C., and Tantillo, D. J. (2013). The viability of nitron-alkene (3+2) cycloadditions in alkaloid biosynthesis. *J. Org. Chem.* 79, 432–435. doi: 10.1021/jo402487d
- Quin, M. B., Flynn, C. M., and Schmidt-Dannert, C. (2014). Traversing the fungal terpenome. *Nat. Prod. Rep.* 31, 1449–1473. doi: 10.1039/C4NP00075G
- Rai, A., Saito, K., and Yamazaki, M. (2017). Integrated omics analysis of specialized metabolism in medicinal plants. *Plant J.* 90, 764–787. doi: 10.1111/tip.13485
- Rai, A., Yamazaki, M., and Saito, K. (2019). A new era in plant functional genomics. *Curr. Opin. Syst. Biol.* 15, 1–10. doi: 10.1016/j.coisb.2019.03.005
- Richter, E., Leaver-Fay, A., Khare, S. D., Bjelic, S., and Baker, D. (2011). De novo enzyme design using Rosetta3. *PLoS One* 6:e19230. doi: 10.1371/journal.pone.0019230
- Ro, D.-K., Paradise, E. M., Ouellet, M., Fisher, K. J., Newman, K. L., Ndungu, J. M., et al. (2006). Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* 440, 940–943. doi: 10.1038/nature04640
- Röthlisberger, D., Khersonsky, O., Wollacott, A. M., Jiang, L., DeChancie, J., Betker, J., et al. (2008). Kemp elimination catalysts by computational enzyme design. *Nature* 453, 190–195. doi: 10.1038/nature06879
- Saito, K., Kobayashi, M., Gong, Z., Tanaka, Y., and Yamazaki, M. (1999). Direct evidence for anthocyanidin synthase as a 2-oxoglutarate-dependent oxygenase: molecular cloning and functional expression of cDNA from a red form of *Perilla frutescens*. *Plant J.* 17, 181–189. doi: 10.1046/j.1365-3113.1999.00365.x
- Sangha, A. K., Davison, B. H., Standaert, R. F., Davis, M. F., Smith, J. C., and Parks, J. M. (2014). Chemical factors that control lignin polymerization. *J. Phys. Chem. B* 118, 164–170. doi: 10.1021/jp411998t
- Sangha, A. K., Parks, J. M., Standaert, R. F., Ziebell, A., Davis, M., and Smith, J. C. (2012). Radical coupling reactions in lignin synthesis: a density functional theory study. *J. Phys. Chem. B* 116, 4760–4768. doi: 10.1021/jp2122449
- Sangha, A. K., Petridis, L., Cheng, X., and Smith, J. C. (2016). Relative binding affinities of monolignols to horseradish peroxidase. *J. Phys. Chem. B* 120, 7635–7640. doi: 10.1021/acs.jpcc.6b00789
- Sato, H., Mitsuhashi, T., Yamazaki, M., Abe, I., and Uchiyama, M. (2018a). Computational studies on biosynthetic carbocation rearrangements leading to quannulatene: initial conformation regulates biosynthetic route, stereochemistry, and skeleton type. *Angew. Chem. Int. Ed.* 57, 14752–14757. doi: 10.1002/anie.201807139
- Sato, H., Narita, K., Minami, A., Yamazaki, M., Wang, C., Suemune, H., et al. (2018b). Theoretical study of sesterfisherol biosynthesis: computational prediction of key amino acid residue in terpene synthase. *Sci. Rep.* 8, 2473–2481. doi: 10.1038/s41598-018-20916-x
- Sato, H., Uchiyama, M., Saito, K., and Yamazaki, M. (2018c). The energetic viability of Δ^1 -piperidine dimerization in lysine-derived alkaloid biosynthesis. *Meta* 8, 48–11. doi: 10.3390/metabo8030048
- Sato, H., Wang, C., Yamazaki, M., Saito, K., and Uchiyama, M. (2018d). Computational study on a puzzle in the biosynthetic pathway of anthocyanin: why is an enzymatic oxidation/reduction process required for a simple tautomerization? *PLoS One* 13, e0198944–e0198910. doi: 10.1371/journal.pone.0198944

- Schlegel, H. B., and Gonzalez, C. (1990). Reaction path following in mass-weighted internal coordinates. *J. Phys. Chem.* 94, 5523–5527. doi: 10.1021/j100377a021
- Siegel, J. B., Zanghellini, A., Lovick, H. M., Kiss, G., Lambert, A. R., St. Clair, J. L., et al. (2010). Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science* 329, 309–313. doi: 10.1126/science.1190239
- Tantillo, D. J. (2010). How an enzyme might accelerate an intramolecular Diels-Alder reaction: theozymes for the formation of salvileucalin B. *Org. Lett.* 12, 1164–1167. doi: 10.1021/ol9028435
- Tantillo, D. J. (2016). Does nature know best? Pericyclic reactions in the *Daphniphyllum* alkaloid-forming cation cascade. *Org. Lett.* 18, 4482–4484. doi: 10.1021/acs.orglett.6b01919
- Tantillo, D. J., Jiangang, C., and Houk, K. N. (1998). Theozymes and compuzymes: theoretical models for biological catalysis. *Curr. Opin. Chem. Biol.* 2, 743–750. doi: 10.1016/S1367-5931(98)80112-9
- Turnbull, J. J., Nakajima, J. I., Welford, R. W. D., Yamazaki, M., Saito, K., and Schofield, C. J. (2004). Mechanistic studies on three 2-oxoglutarate-dependent oxygenases of flavonoid biosynthesis: anthocyanidin synthase, flavonol synthase, and flavanone 3 β -hydroxylase. *J. Biol. Chem.* 279, 1206–1216. doi: 10.1074/jbc.M309228200
- Ujaque, G., Tantillo, D. J., Hu, Y., Houk, K. N., Hotta, K., and Hilvert, D. (2002). Catalysis on the coastline: theozyme, molecular dynamics, and free energy perturbation analysis of antibody 21D8 catalysis of the decarboxylation of 5-nitro-3-carboxybenzisoxazole. *J. Comput. Chem.* 24, 98–110. doi: 10.1002/jcc.10151
- Wall, M. E. (1998). Camptothecin and taxol: discovery to clinic. *Med. Res. Rev.* 18, 299–314. doi: 10.1002/(SICI)1098-1128(199809)18:5<299::AID-MED2>3.0.CO;2-O
- Wall, M. E., Wani, M. C., Cook, C. E., Palmer, K. H., McPhail, A. T., and Sim, G. A. (1966). Plant antitumor agents. I. The isolation and structure of camptothecin, a novel alkaloidal leukemia and tumor inhibitor from *Camptotheca acuminata* 1,2. *J. Am. Chem. Soc.* 88, 3888–3890. doi: 10.1021/ja00968a057
- Wu, J., and Liu, L. F. (1997). Processing of topoisomerase I cleavable complexes into DNA damage by transcription. *Nucleic Acids Res.* 25, 4181–4186. doi: 10.1093/nar/25.21.4181
- Yang, T., Nagy, I., Mancinotti, D., Otterbach, S. L., Andersen, T. B., Motawia, M. S., et al. (2017). Transcript profiling of a bitter variety of narrow-leaved lupin to discover alkaloid biosynthetic genes. *J. Exp. Bot.* 68, 5527–5537. doi: 10.1093/jxb/erx362
- Ye, Y., Minami, A., Mandi, A., Liu, C., Taniguchi, T., Kuzuyama, T., et al. (2015). Genome mining for sesterterpenes using bifunctional terpene synthases reveals a unified intermediate of di/sesterterpenes. *J. Am. Chem. Soc.* 137, 11846–11853. doi: 10.1021/jacs.5b08319
- Zou, H.-Y., Zhao, D.-X., and Yang, Z.-Z. (2013). A theoretical study on mechanism of the anticancer drug camptothecin's E-ring-opening. *J. Mol. Graph. Model.* 43, 58–65. doi: 10.1016/j.jmgm.2013.03.004

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Sato, Saito and Yamazaki. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Large-Scale Profiling of Saponins in Different Ecotypes of *Medicago truncatula*

Zhentian Lei^{1,2*}, Bonnie S. Watson³, David Huhman³, Dong Sik Yang^{3,4} and Lloyd W. Sumner^{1,2*}

¹ University of Missouri Metabolomics Center, Columbia, MO, United States, ² Department of Biochemistry, University of Missouri, Columbia, MO, United States, ³ Noble Research Institute, Ardmore, OK, United States, ⁴ Biomaterials Laboratory, Material Research Center, Samsung Advanced Institute of Technology, Gyeonggi-do, South Korea

OPEN ACCESS

Edited by:

Kazuki Saito,
RIKEN Center for Sustainable
Resource Science (CSRS), Japan

Reviewed by:

Jacob Pollier,
Flanders Institute for Biotechnology,
Belgium
Ery Odette Fukushima,
Regional College Amazon Ikiam,
Ecuador

Cristian Daniel Quiroz Moreno,
Regional College Amazon Ikiam,
Ecuador, in collaboration with
reviewer EF

*Correspondence:

Zhentian Lei
leiz@missouri.edu
Lloyd W. Sumner
sumnerlw@missouri.edu

Specialty section:

This article was submitted to
Plant Metabolism
and Chemodiversity,
a section of the journal
Frontiers in Plant Science

Received: 16 April 2019

Accepted: 13 June 2019

Published: 03 July 2019

Citation:

Lei Z, Watson BS, Huhman D,
Yang DS and Sumner LW (2019)
Large-Scale Profiling of Saponins
in Different Ecotypes of *Medicago*
truncatula. *Front. Plant Sci.* 10:850.
doi: 10.3389/fpls.2019.00850

A total of 1,622 samples representing 201 *Medicago truncatula* ecotypes were analyzed using ultrahigh pressure liquid chromatography coupled to mass spectrometry (UHPLC-MS) to ascertain saponin profiles in different *M. truncatula* ecotypes and to provide data for a genome-wide association study and subsequent line selection for saponin biosynthesis. These ecotypes originated from 14 different Mediterranean countries, i.e., Algeria, Cyprus, France, Greece, Israel, Italy, Jordan, Libya, Morocco, Portugal, Spain, Syria, Tunisia, and Turkey. The results revealed significant differences in the saponin content among the ecotypes. European ecotypes generally contained higher saponin content than African ecotypes ($p < 0.0001$). This suggests that *M. truncatula* ecotypes modulate their secondary metabolism to adapt to their environments. Significant differences in saponin accumulation were also observed between the aerial and the root tissues of the same ecotypes ($p < 0.0001$). While some saponins were found to be present in both the aerial and root tissues, zanhic acid glycosides were found predominantly in the aerial tissues. Bayogenin and hederagenin glycosides were found mostly in roots. The differential spatially resolved accumulation of saponins suggests that saponins in the aerial and root tissues play different roles in plant fitness. Aerial saponins such as zanhic glycosides may act as animal feeding deterrent and root saponins may protect against soil microbes.

Keywords: *Medicago truncatula*, ecotypes, triterpene saponin, metabolomics, LC-MS/MS

INTRODUCTION

Legumes are economically important and widely cultivated crops. They contain relatively high protein content and are important sources of protein for both humans and animals. Their high protein content may be attributed to their unique symbiotic relationship with nitrogen-fixing bacteria. Legumes also produce a vast array of natural products including flavonoids, isoflavonoids, anthocyanins, condensed tannins, lignin, and saponins (Dixon and Sumner, 2003). These natural products play important roles in many important biological processes and are important to legume quality. For example, flavonoids serve as signaling compounds in the symbiotic plant-microbe interactions and induce the expressions of Nod genes in the nitrogen-fixing bacteria. Condensed tannins can prevent bloat associated with animals grazing on legumes with high protein

content such as alfalfa and clover. Saponins are triterpene glycosides composed of triterpenoid aglycones (normally referred to as sapogenins) conjugated with various carbohydrate residues. They have been documented to possess anti-fungal, anti-bacterial and anti-insect properties and contribute to plant development and defense against pathogens (Moses et al., 2014). However, due to their hemolytic activity and membrane permeabilization nature, saponins are considered as anti-nutritional. They have been reported to cause bloat, reduce digestibility of proteins, interfere with uptake of nutrients in the gut, and result in reduced weight gain (Francis et al., 2002). These undesired anti-nutritional effects of saponins negatively affect the efficient use of high-protein containing legume forages such as alfalfa and clover as animal feeds. Manipulation of the saponin contents in legumes through genetic engineering and/or molecular breeding may provide an efficient way to improve the nutritional values or field performance of legume forages. However, this effort is hindered by our limited understanding of triterpene saponin biosynthesis. In addition, the effects of growth conditions and developmental stages on levels of individual saponins in legumes are still not clear. Information about saponin variation among the many different ecotypes is also lacking. This knowledge is particularly useful in breeding of low saponin containing legume forages.

LC-MS based metabolomics is ideally suited for the analysis of saponins in complex plant extracts. It has been successfully used in the analyses of saponins in many legumes including *Medicago truncatula* (Huhman et al., 2005; Kapusta et al., 2005a,b), *M. arborea* (Tava et al., 2005), alfalfa (Sen et al., 1998; Bialy et al., 1999), clover (Perez et al., 2013), and soybean. Using LC-Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS), Pollier et al. (2011) revealed a complex mixture of saponins in the hairy roots of *M. truncatula*. More recently, saponins in 12 annual *Medicago* species have been profiled and compared (Tava and Pecetti, 2012). The content of saponins were found to range from 0.38 to 1.35% (dry weight), depending on the species. In addition, differences in the aglycone moieties were observed among the 12 *Medicago* species. While some aglycones such as bayogenin and hederagenin were found in all the species, some including medicagenic acid and zanhic acid were species-dependent (Tava and Pecetti, 2012). The large number of MS/MS and NMR data collectively generated over the past years by a number of different groups constitutes an important and valuable resource for saponin annotation in *M. truncatula* (Bialy et al., 1999; Kapusta et al., 2005a,b; Tava et al., 2005; Pollier et al., 2011).

Saponins are broadly classified into hemolytic (oleanates) and non-hemolytic (soyasapogenol) (Figure 1). It is generally believed that the hemolytic activity of oleanate saponins is conferred by the presence of a C28-carboxylic group (Voutquenne et al., 2002). Glycosylation of the C28-carboxylic group dramatically reduces and even eliminates the hemolytic activity. Our understanding of triterpene saponin biosynthesis in *M. truncatula* is still very limited. All the saponins are believed to derive from beta-amyrin that is formed through the cyclization of 2,3-oxidosqualene catalyzed by beta-amyrin synthase (Figure 1) (Tava et al., 2011). Hydroxylation and subsequent oxidation of beta-amyrin lead to multiple pentacyclic triterpene aglycones (or sapogenins), glycosylation of which

results in diverse and complex saponins. The hydroxylation and oxidation of aglycones are believed to be catalyzed by cytochrome P450 proteins (Tava et al., 2011). In *M. truncatula*, a well-established model legume and close relative to alfalfa (*Medicago sativa*), CYP716A12 has been identified to oxidize beta-amyrin to erythrodiol and then subsequently to oleanolic acid (Carelli et al., 2011; Fukushima et al., 2013). Mutants in CYP716A12, *lha* (lacking hemolytic activity), were found to lack hemolytic saponins and only produce non-hemolytic soyasapogenins (Carelli et al., 2011). Oxidation of oleanolic acid to hederagenin, gypsogenin, and gypsogenic acid was found to be catalyzed by CYP72A68 (Tzin et al., 2019). Formation of non-hemolytic triterpene aglycones 24-hydroxy-beta-amyrin and soyasapogenol B from beta-amyrin is catalyzed subsequently by CYP93E2 and CYP72A61v2 (Fukushima et al., 2013). A cytochrome P450 (CYP72A67) involved in hemolytic sapogenin biosynthesis was also identified. It was found to be responsible for hydroxylation at the C-2 position of oleanolic acid for downstream sapogenin biosynthesis (Biazzi et al., 2015). However, enzymes involved in the formation of other sapogenins are still unknown. In addition, the effects of the environment on the production of saponins in *M. truncatula* is not clear. To increase our understanding of saponin accumulation in different ecotypes and provide a basis for selecting appropriate *M. truncatula* lines for future correlated gene expression analyses and for the discovery of genes involved in triterpene aglycone biosynthesis, we profiled saponins in 1,622 samples representing 201 *M. truncatula* ecotypes from 14 different countries using UHPLC-MS (Table 1). Differential distributions of saponins in roots and aerial tissues were observed. Zanhic acid saponins were found predominantly in leaves, whereas hederagenin and bayogenin saponins were mostly found in roots. The differential spatial accumulation suggests that different classes of saponins may play different roles in plant defense responses. In addition, different ecotypes were found to accumulate different amounts of saponins, with highest saponin containing ecotypes found mostly in Europe and lowest saponin containing ecotypes were from Africa.

MATERIALS AND METHODS

Biological Materials

Seeds were scarified with sulfuric acid, sterilized with 5% bleach and germinated on damp sterile filter paper. Three days after germination seedlings were transplanted into root cones (Stuewe & Sons Inc.) filled with Turface (BWI Texarkana) which had been rinsed in distilled water and autoclaved. The ecotypes were grown in a growth room under controlled conditions. Day length was 16 h, with a gradual increase in light. The light source included both fluorescent and incandescent bulbs, and light intensity averaged ~225 μ J. Temperature was set at a constant 21°C. Plants were watered with Broughton and Dilworth (100 ppm N) fertilizer every day. Enough seedlings were planted to analyze four replicates of each ecotype. In most cases, this was accomplished but a few ecotypes grew poorly and three or fewer replicates were harvested. The large number of plants necessitated growth in three separate groups, and A17 and R108

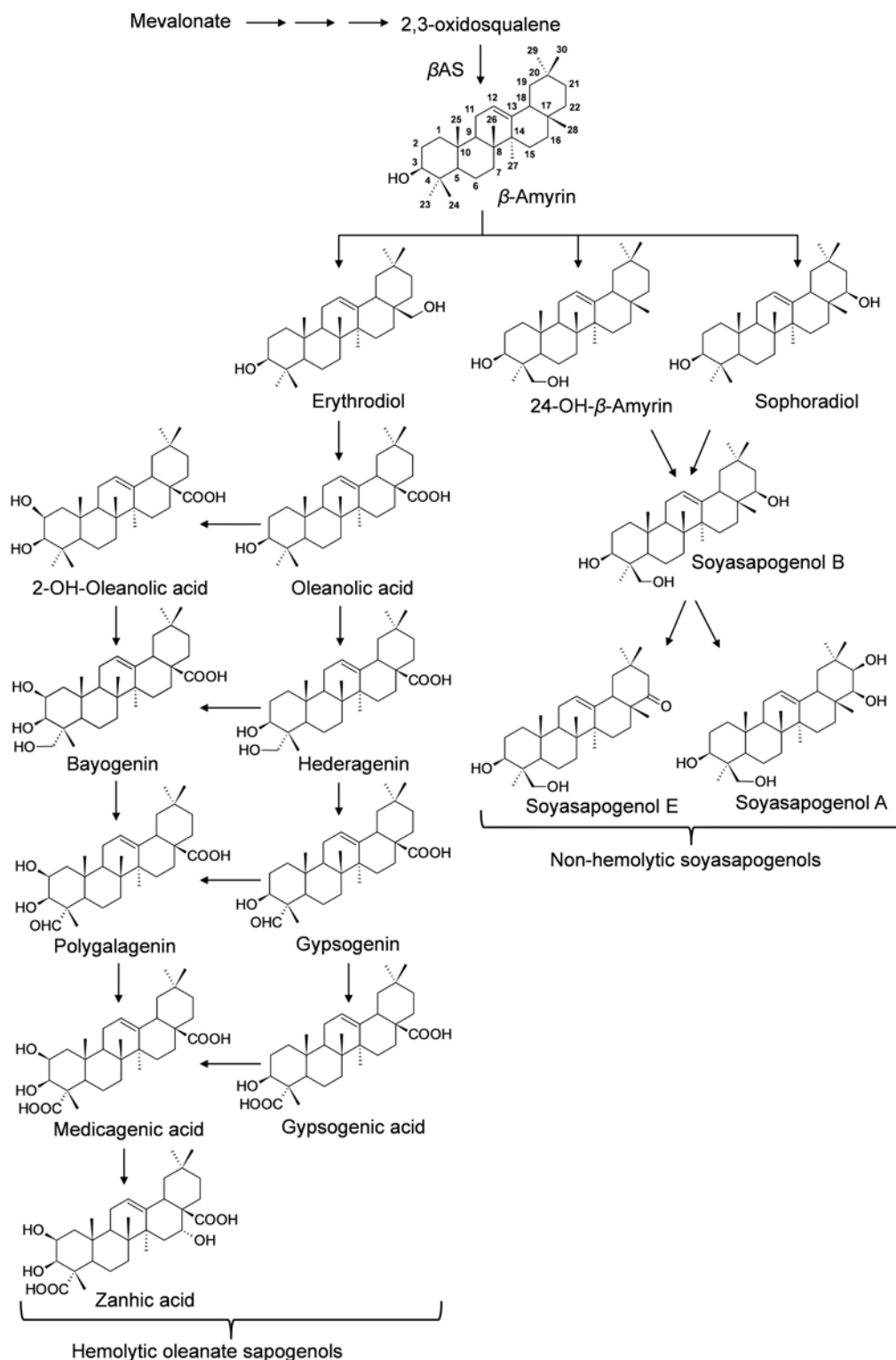


FIGURE 1 | Proposed Biosynthesis of Triterpene Sapogenins in *M. truncatula*. All the sapogenins are derived from beta-amyrin through a series of oxidations that include hydroxylation and subsequent oxidations catalyzed by P450s. Glycosylation of the sapogenins results in the diversity of saponins. In *M. truncatula*, CYP716A12 has been identified to oxidize beta-amyrin to erythrodiol and subsequently to oleanolic acid, and CYP72A68 was found to convert to oleanolic acid, gypsogenin, and subsequently to gypsogenic acid (Fukushima et al., 2013; Tzin et al., 2019). CYP93E2 oxidize beta-amyrin to 24-hydroxy-beta amyrin which is converted to soyasapogenol B by CYP72A61v2 (Fukushima et al., 2013). CYP72A67 responsible for hydroxylation at the C-2 position of oleanolic acid for downstream sapogenin biosynthesis was also identified (Biazzi et al., 2015).

TABLE 1 | Geographical origin of the 201 *M. truncatula* inbred lines used for saponin profiling.

Geographic origin	Number of lines
North Africa	
Algeria	41
Libya	8
Morocco	10
Tunisia	14
Europe	
France	28(13) ^a
Greece	16(2) ^b
Italy	9
Portugal	11(1) ^c
Spain	32
Middle East	
Cyprus	9
Israel	5
Jordan	2
Syria	2
Turkey	3
Others of unknown origin	11

^a13 out of the 28 lines were from Corsica. ^b2 out of the 16 lines were from Crete.

^c1 out of the 11 lines was from Madeira.

were grown with each group as controls. Plantings were staggered so that all plants could be harvested during the same morning time frame (3–5 h after full light) to insure a uniform position in the diurnal photosynthetic cycle. Plants were harvested at 5 weeks of age before the onset of flowering. Roots and aerial tissues were separated and immediately frozen in liquid nitrogen.

The tissues were lyophilized and dry weights were recorded. All tissues were ground to a fine powder and 10 mg each was weighed for extraction. If plant material was limited, extraction volumes were reduced proportionally. Tissues were incubated with 80:20 methanol:water solution containing 18 µg/ml umbelliferone.

UHPLC-QTOF-MS

Roots and aerial tissues were lyophilized until dry and ground to a fine powder. Ten milligrams (10 ± 0.06 mg) of powder for all tissues were accurately weighed and extracted on an orbital shaker for 2 h with 1 mL of 80% of methanol containing 18 µg/mL umbelliferone as an internal standard. Samples were centrifuged at $2,900 \times g$ for 30 min at 4°C, and the supernatants were collected. Five microliters of the supernatant were injected into a Waters UPLC system coupled to a quadrupole-time of flight mass spectrometer (QTOF-MS, Waters QTOF Premier). Chromatographic separations were performed on a Waters reverse phase column (2.1×150 mm, BEH C18, 1.7 µm particles) using the following gradient: mobile phase B (acetonitrile) increased from 5 to 70% over 30 min, then to 95% in 3 min, held at 95% for 3 min, and returned to 95% mobile phase A (0.05% formic acid in water) for equilibration for 3 min. The flow rate of the mobile phases was 0.56 mL/min, and the column and autosampler temperatures were maintained at 60

and 4°C, respectively. Mass spectral data were acquired from m/z 50 to 2,000 in the negative electrospray ionization mode, with the nebulization gas set at 850 L/h (350°C) and the cone gas at 50 L/h (120°C). Raffinose (m/z 503.1612) was used as the reference compound in the independent lock-mass mode, with the lock mass scan (1 s) collected every 10 s for accurate mass measurements. The concentration of raffinose was 50 fmol/mL, and the flow rate 0.2 mL/h.

The raw data files obtained from UHPLC-QTOF-MS analyses were processed with MarkerLynx software (version 4.1, Waters) for mass features extraction and alignment with the following parameters: minimum peak intensity: 500 counts, mass tolerance: 0.05 Da, and retention time window 0.2 min. The peak areas were normalized by dividing each peak area by the value of the internal standard peak (area of metabolite/area of area of internal standard $\times 1,000$). Annotations of metabolites were performed by matching their m/z to those of the previously observed saponins in *M. truncatula* (Huhman et al., 2005; Kapusta et al., 2005a,b; Pollier et al., 2011). Retention time was also used in the Rt- m/z pair matching when it was available. Tandem MS was performed on a number of saponins, mainly the most abundant ones, to validate the annotations by matching to previously reported MS/MS data. The MS/MS experiments were performed using a UHPLC-Bruker QTOF MS. The tandem data were compared to previously published data for identification confirmation (Kapusta et al., 2005a; Pollier et al., 2011). The normalized data (i.e., data normalized to the internal standard) were used for statistical analyses. Multivariate statistical analyses were performed using JMP software from SAS (Cary, NC). Tukey HSD (Tukey Honest Significant Differences) was performed using TukeyHSD function in R, and the not significantly different groups were labeled with the same letters using the HSD.test function in the “agricolae” package.

RESULTS AND DISCUSSION

The *M. truncatula* ecotypes (201 lines, 1,622 plants) analyzed in this work represented 14 geographical origins, i.e., Algeria, Cyprus, France, Greece, Israel, Italy, Jordan, Libya, Morocco, Portugal, Spain, Syria, Tunisia, and Turkey (Table 1). The raw data are freely available for download at <https://sumnerlab.missouri.edu/download/>. Aerial and root tissues were separated, lyophilized and weighed. The dry weight for each ecotype’s aerial and root tissues is shown in Figure 2. Almost all the ecotypes (93.3%) were found to produce more aerial tissues than roots by weight. Significant difference in the average total dry weight was observed among the ecotypes ($p < 0.0001$), with a 90 fold difference between the lowest line (HM095, France, 6.42 ± 9.43 mg, mean \pm standard deviation, $n = 5$) and the highest line (HM174, Spain, 548.98 ± 68.25 mg, mean \pm standard deviation, $n = 5$). It was also found that there was a significant positive correlation between the aerial and the root dry weight ($r = 0.86$, $p < 0.0001$), indicating that ecotypes producing more aerial tissues also tended to produce more root tissues. The average total dry weight for ecotypes of the same country of origin is shown in Table 2. Most of the North African

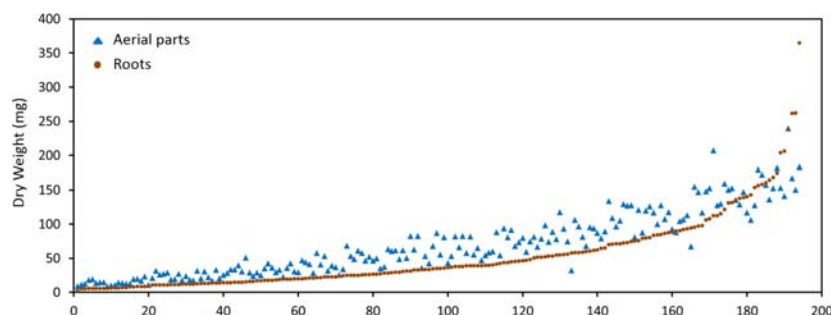


FIGURE 2 | Dry weight distribution of the *M. truncatula* ecotypes. A great majority of the ecotypes (92.3%) were found to produce more aerial tissues (closed triangles) than roots (closed squares) by weight. Significant difference in the dry weight was observed among the ecotypes ($p < 0.0001$) with the highest total dry weight being 548.98 ± 68.25 mg (HM174, Spain) and the lowest 6.42 ± 9.43 mg (HM095, France). Significant positive correlation between the aerial and the root dry weight was also observed ($r = 0.86$, $p < 0.0001$), indicating that ecotypes that produced more aerial tissues also tended to produce more root tissues.

and Middle East ecotypes appeared to produce less biomass compared to the European ecotypes, but the difference was not statistically significant ($p = 0.3$). This is due to large variations of dry weight among the ecotypes from the same country as evidenced by the large standard deviation associated with each average dry weight (Table 2). Thus, the dry weight of ecotypes, when grown in greenhouse, appeared to be ecotype specific and did not show a statistically significant correlation between their dry weight and their geographic origin. This is probably due to the different growth rates of the different ecotypes. For example, the two well characterized and mostly used ecotypes, A17 and R108, have previously been found to have different seed-to-seed generation times (Hoffmann et al., 1997). The generation time of R108 was 12–14 weeks, which is about 3 weeks shorter than A17. Ecotypes with even shorter seed generation times (e.g., 7 weeks shorter than A17) was also found (Hoffmann et al., 1997). As the ecotypes were harvested at the same time and not at the same developmental stage, the different generation time among the ecotypes may contribute to the high variance observed (Table 2). Although it was desirable to harvest all the ecotypes at the same developmental stage, it was simply not feasible given the scale of this experiment.

Saponin profiling of the aerial and root tissues was performed in a random manner and spanned over a period of 5 months (November 2012–March 2013). The data are shown in Supplementary Tables 1,2. Reproducibility is always a concern when large-scale experiments of this magnitude are performed. To monitor the reproducibility, a blank wash and a quality control (QC) mixture were performed every 10 samples to monitor any carry-over or changes in instrumental response. An internal standard solution (extraction solution only) was also analyzed every 20 samples. Eighty-seven injections of the internal standard solution were made during the 5-months of analyses and the responses were used to calculate the relative standard deviation (RSD) to quantify reproducibility. The RSD was determined to be 15.3%, comparable to a previously reported value (15.9%) in a metabolomics project (Kirwan et al., 2013) and below that (20%) recommended for large-scale metabolomics (Food and Drug Administration [FDA], 2001; Zelena et al., 2009). Annotation of saponins was performed by matching the

mass features' m/z to those of saponins found in *M. truncatula* and then confirmed by MS/MS (Figure 3). Figure 3 shows a representative UHPLC-MS chromatogram of the aerial and root tissues of HapMap 135 (line number: L000332, country of origin: Israel). Significant differences were observed between the metabolic profiles of the aerial and root tissues (Figures 3A,B). For example, a peak (m/z 973.5069_Rt12.48 min) found predominantly in the root tissues was annotated as glucose-glucose-glucose-bayogenin and confirmed by tandem MS (Figures 3C,D). The relative abundances of saponins (area of metabolite peak normalized to that of the internal standard in the sample) were used for statistical analysis such as principal component analysis (PCA) and multiple mean comparisons. The results showed that there was significant difference in the saponin content among different ecotypes ($p < 0.0001$) as well as between the aerial and the root tissues of the same ecotypes ($p < 0.0001$). These data indicated that the geographical origin has an impact on the production of saponin in plants and that saponins differentially accumulated in roots and aerial tissues. Their differences and biological implications are discussed below.

TABLE 2 | Average total dry weights and their standard deviations of ecotypes.

Country of origin	Average aerial dry weight (mg/line)	Standard error	Average root dry weight (mg/line)	Standard error
Algeria	61.21	6.45	45.98	7.37
Cyprus	50.52	11.80	22.87	4.81
France	91.93	11.89	81.11	14.31
Greece	77.74	10.83	50.11	8.89
Israel	104.72	36.14	87.20	38.89
Italy	91.79	13.59	62.93	17.87
Jordan	57.66	17.53	18.20	5.73
Libya	69.71	9.74	42.79	6.35
Morocco	92.91	13.07	62.68	9.35
Portugal	81.28	10.80	56.74	16.68
Spain	77.23	9.59	62.67	12.40
Syria	51.49	31.00	25.63	12.68
Tunisia	68.26	13.62	52.48	13.29
Turkey	71.74	10.42	28.82	6.80

The average total dry weight was calculated by averaging the average weights of all individual *M. truncatula* lines from the same country of origin.

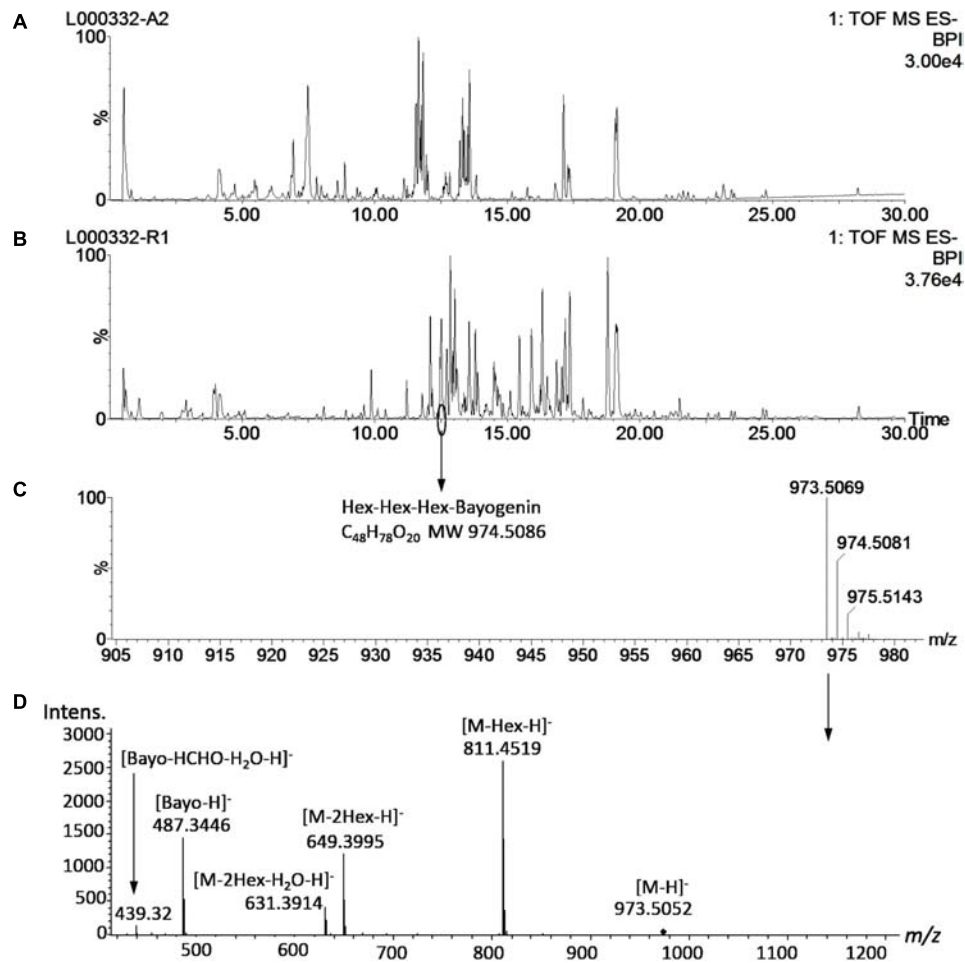
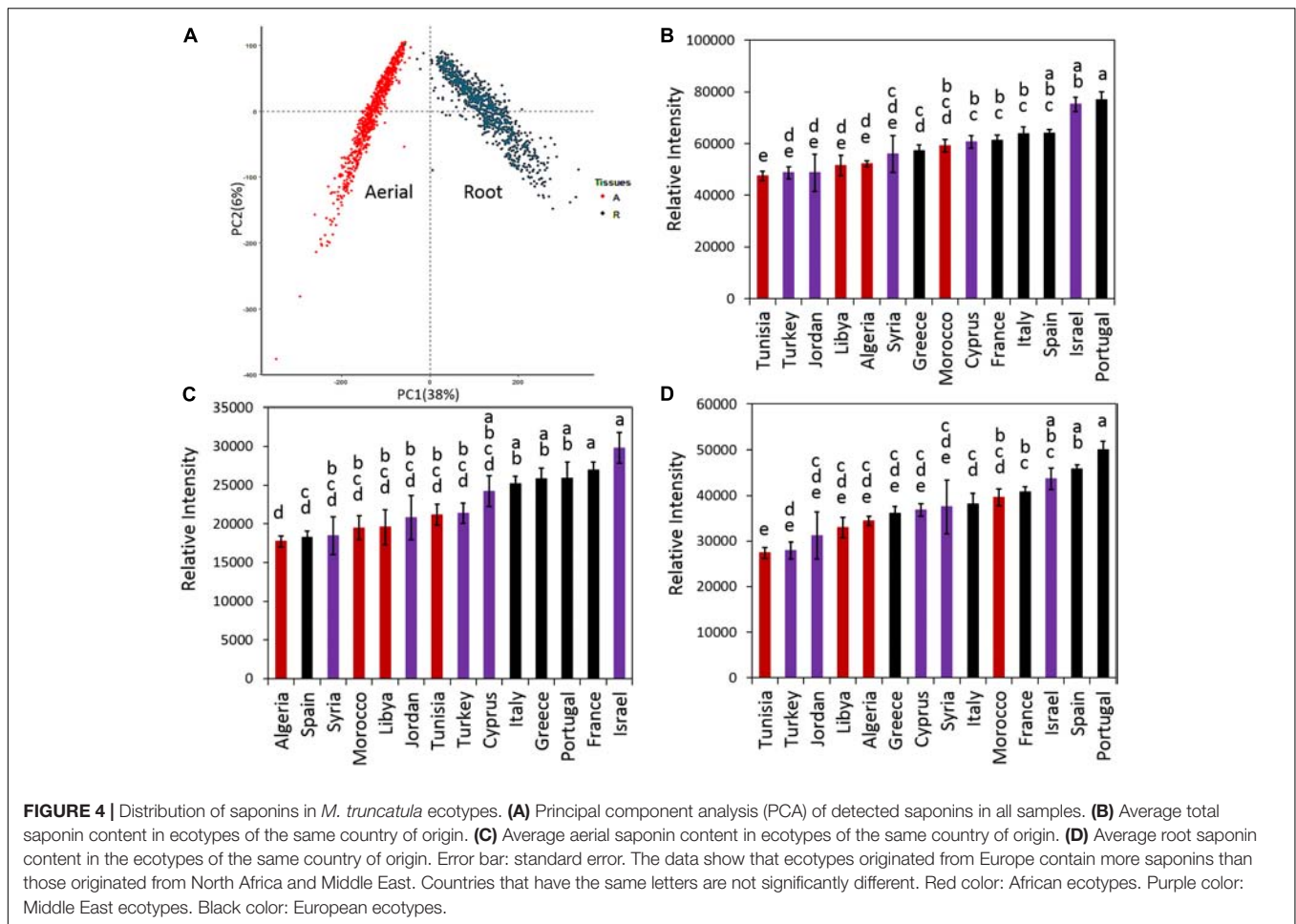


FIGURE 3 | Representative UHPLC-MS metabolite profiles of *M. truncatula*. **(A)** UHPLC-ESI-MS chromatogram of aerial tissue extract of *M. truncatula* HapMap 135 (line number: L000332, country of origin: Israel). **(B)** UHPLC-ESI-MS chromatogram of root tissue extract of HapMap 135. **(C)** Mass spectrum of a peak at Rt 12.48 min in the root tissue extract showing m/z at 973.5183 and identified as hexose-hexose-hexose-bayogenin. **(D)** MS/MS spectrum of the precursor ion at m/z 973.5 at a collision energy ramped from 55 to 90 V, confirming glucose-glucose-glucose-bayogenin. MS spectrum was acquired with a QTOF Premier (Waters) and MS/MS spectrum with a MaXis Impact QTOF (Bruker Daltonics). The data show that hexose-hexose-hexose-bayogenin is predominantly accumulated in the roots of *M. truncatula*. Bayo: bayogenin; Hex: hexose.

Saponin Content Among Ecotypes

Principal component analysis was performed for all samples and means for the total, aerial and root saponin content in ecotypes of the same country of origin were calculated (Figure 4). Figure 4 shows the PCA results (Figure 4A), mean of total saponins (Figure 4B), mean of aerial saponins (Figure 4C) and mean of root saponins (Figure 4D) in ecotypes from the same country. There was a significant difference in the total saponin content among the ecotypes of different countries ($p < 0.0001$). The lowest saponin-containing ecotypes were found to originate from Tunisia and contained only 60% of the amount of saponins in the highest saponin-containing ecotypes (Portugal). Figure 4B also shows that the low saponin-containing ecotypes were mostly from Africa except those originating from Turkey. These ecotypes (Algeria, Libya, Tunisia, and Turkey) contained comparable amounts of total saponins ($p = 0.4$). This may be explained by the similar environments

such as similar climates, seasonal changes and soil conditions in these countries as most of them are in the African continent. In contrast, ecotypes originating from Israel and European countries such as France, Italy, Portugal, and Spain typically contained higher amount of saponins. The two highest saponin-containing ecotypes were from Portugal and Israel. This distinct difference between the African and the European ecotypes indicated a clear geographic segregation of *M. truncatula* around the Mediterranean area in terms of saponin production, with the ecotypes in the north region (Europe) segregated from those in the south region (Africa). Similar segregation has also been observed in a previous microsatellite diversity study of 346 inbred lines of *M. truncatula* ecotypes that revealed a stratification of the *M. truncatula* population between the North and the South of the Mediterranean basin (Ronfort et al., 2006). It was further suggested that the *M. truncatula* colonization of the Mediterranean region was via two routes from its original



habitat around the Middle East (Ronfort et al., 2006). However, the molecular mechanism responsible for the difference, i.e., higher saponin contents in European ecotypes and lower saponin contents in the African ecotypes, is not clear. Plant secondary metabolism is complex and influenced by both biotic and abiotic stimuli. Plants under different environments can modulate their secondary metabolism to increase their fitness and adapt to the environments. The effect of environments on plant saponin biosynthesis has been recently reviewed (Szakiel et al., 2011). Both biotic and abiotic factors (e.g., temperature, light, humidity, water availability, soil fertility, insects, herbivores, competition from neighboring plants) and their interactions all can affect saponin biosynthesis. For example, saponin content was found higher in aphid-infested alfalfa compared to the uninfested alfalfa (Goławska et al., 2012). A study of herbivore-induced responses in alfalfa further indicated that saponin content increased with higher herbivore densities (Agrell et al., 2003). Methyl jasmonate treatment that mimics mechanical wounding of plants was also found to increase the production of saponins in *M. truncatula* (Suzuki et al., 2002), suggesting that grazing also results in higher saponin content in forage legumes. The environmental abiotic factors also affect saponin content significantly. It has been shown that saponin content decreased significantly in medicinal plants exposed to drought stress (Solíz-Guerrero et al., 2002;

Zhu et al., 2009) and the application of an appropriate amount of inorganic fertilizer was able to partly restore saponin content in *Bupleurum* (Zhu et al., 2009). In *Brachiaria*, the main forage for ruminants cultivated worldwide in both tropical and subtropical climates, the saponin content was found to correlate negatively with the duration of sunshine and maximum ambient temperature, but positively with relative humidity (Lima et al., 2012). This suggests that the combination of high temperature, long duration of sunshine and arid condition in Africa might, at least in part, be responsible for the lower saponin content in the African ecotypes. A recent study using over 20,000 annotated genes from *M. truncatula* showed that genes involved in defense against pathogens and herbivores constituted the single largest functional class of genes under positive selection in adaptive evolution (Paape et al., 2013). Other genes under positive selection included those involved in mediating symbiotic relationship with rhizobia and one-third of the annotated histone-lysine methyltransferases that could be involved in epigenetic modifications (Paape et al., 2013). The relatively higher saponin content in European ecotypes may reflect such a positive selection of genes related to disease and defense response as saponins are anti-feeding and anti-microbial (Da Silva et al., 2012; Goławska et al., 2012). Significant variations in the response of *M. truncatula* ecotypes to *Verticillium albo-atrum*, a soil-borne

pathogenic fungus, have been reported recently (Ben et al., 2013). Comparison of the resistant and susceptible *M. truncatula* ecotypes led to the identification of three QTLs associated with resistance to the *Verticillium* wilt in the resistant lines. The resistance appeared to be selected within environments as it did not seem to correlate with the population structure (Ben et al., 2013).

In addition to the total saponin content, the aerial and root tissues of the African ecotypes were also found to contain significantly less saponins than the European ecotypes ($p < 0.0001$) (Figures 4C,D). In the aerial tissues, the lowest saponin containing ecotypes were found to be from Algeria. They contained only 60% of the saponins found in the highest saponin-containing ecotypes (Israel). In roots, Tunisia's ecotypes contained the least amount of saponin, only 55% of the saponin in the highest saponin-containing ecotypes (Portugal). Figure 4 also reveals that roots contained higher saponin content than the aerial parts. For example, compared to their aerial parts, the roots of Portugal's ecotypes contained about twice the amount of saponins (Figures 4C,D). The differences between the aerial and the root tissues were not only quantitative but also qualitative as evidenced by the distinct saponin profiles in the aerial parts (Figure 3A) and the roots (Figure 3B). This was further supported by PCA analysis that showed the aerial tissues were clearly segregated from the root tissues (Figure 4A), suggesting that the difference between the aerial and the root tissues would be greater than the difference among ecotypes. The differential accumulation of saponins in the aerial and the root tissues may reflect the different roles of these saponins in plant fitness and defense response as discussed below.

Differential Accumulation of Saponins in the Aerial and Root Tissues

The results of the saponin profiling in the aerial and the root tissues are shown in Figure 5. Figure 5 shows the accumulation of individual classes of saponins (i.e., bayogenin glycosides, hederagenin glycosides, medicagenic acid glycosides, soyasaponin B glycosides, soyasaponin E glycosides, and zanhic acid glycosides) in the aerial (Y-axis) and the root (X-axis) tissues of the ecotypes. It reveals clear differences in the spatial accumulation of saponins. Regardless of the country of origin, bayogenin glycosides, hederagenin glycosides and soyasapogenin E glycosides were mostly found in the roots, while zanhic acid glycosides were only detected in the aerial tissues. In contrast, medicagenic acid glycosides and soyasaponin B glycosides were found in both the aerial and root tissues although the aerial tissues appeared to contain more medicagenic acid glycosides. The finding is similar to a previous report that zanhic acid glycosides, medicagenic acid glycosides and soyasaponin B glycosides were found to be the three dominant groups of saponins in *M. truncatula* foliar tissues (Kapusta et al., 2005b; Confalonieri et al., 2009). It is also consistent with several previous reports that zanhic acid was the major aglycone in the aerial parts of *M. truncatula* (Kapusta et al., 2005a,b) and zanhic acid glycosides could not be detected in the roots of *M. truncatula* (Confalonieri et al., 2009). A more recent study of

M. truncatula hairy roots showed that the zanhic acid glycosides were also absent in the hairy roots (Pollier et al., 2011). Zanhic acid glycosides are therefore leaf-specific saponins. This suggests that P450 enzyme(s) involved in the biosynthesis of zanhic acid can best be studied using leaf tissues. Comparisons of P450 gene expression profiles between the aerial and root tissues may facilitate the identification of enzymes converting medicagenic acid into zanhic acid. Zanhic acid formation is the last step of the sapogenol biosynthesis pathway (Figure 1). The accumulation of zanhic acid glycosides and the lack of other saponins in the earlier steps of the pathway in the aerial tissues suggest that the physiological role of zanhic acid glycosides cannot be substituted for by other saponins such as hederagenin and bayogenin glycosides. The differential accumulation of saponins in the aerial and the root tissues may reflect the effect of the different environments on the aerial parts and the roots. Compared to roots, one of the unique stresses that the aerial parts face is herbivore feeding and wounding. This suggests that zanhic acid glycosides may be potent anti-feeding metabolites. Indeed, some zanhic acid glycosides have been found to be the most bitter and throat-irritating components among the complex saponins in alfalfa (Oleszek et al., 1992) and the most active compounds in disrupting the transmural potential difference in mammalian small intestine (Oleszek et al., 1994). Therefore, zanhic acid glycosides are generally considered as anti-nutritional and the major anti-feeding agents against herbivores in leaves. Similar to alfalfa, medicagenic acid glycosides were also found in the aerial parts of *M. truncatula* ecotypes and their amounts were higher in the aerial parts than in the roots (Figure 5). Unlike zanhic acid glycosides, medicagenic acid glycosides have been reported to possess, in addition to the anti-feeding property, a broad and strong anti-microbial activity (Oleszek et al., 1990; Oleszek, 1996; Avato et al., 2006; Jarecka et al., 2008). For example, medicagenic acid glycosides strongly inhibited the growth of *Trichoderma viride*, a fungus that is highly sensitive to alfalfa saponins and has been traditionally used to quantify saponins (Zimmer et al., 1967). In contrast, zanhic acid glycosides were found inactive against a wide range of fungi including *T. viride*; even at higher concentrations (Oleszek et al., 1992; Oleszek, 1996). Because the difference between medicagenic acid and zanhic acid is the presence of C16-hydroxy group in zanhic acid, it has been suggested that the C16-hydroxy group is responsible for the strong bitter taste but low anti-fungal activity of zanhic acid glycosides (Oleszek, 1996). Soyasaponin B glycosides were also found to accumulate in the aerial parts and the roots (Figure 5). This is consistent with the previous report that soyasaponin B glycosides were found in both aerial and root tissues of alfalfa (Sen et al., 1998). Soyasaponin B glycosides are non-hemolytic saponins. They have been reported to possess anti-feeding and antifungal activities. The combination of saponins of distinct functions in leaves provides an excellent defense against herbivores and fungal attacks. Indeed, incorporation of dried alfalfa leaf tissue in their diet significantly inhibited growth and development of larvae of the European corn borer (*Ostrinia nubilalis*). In contrast, saponin fractions isolated from alfalfa root tissues, when incorporated into their diet at equivalent concentrations, had

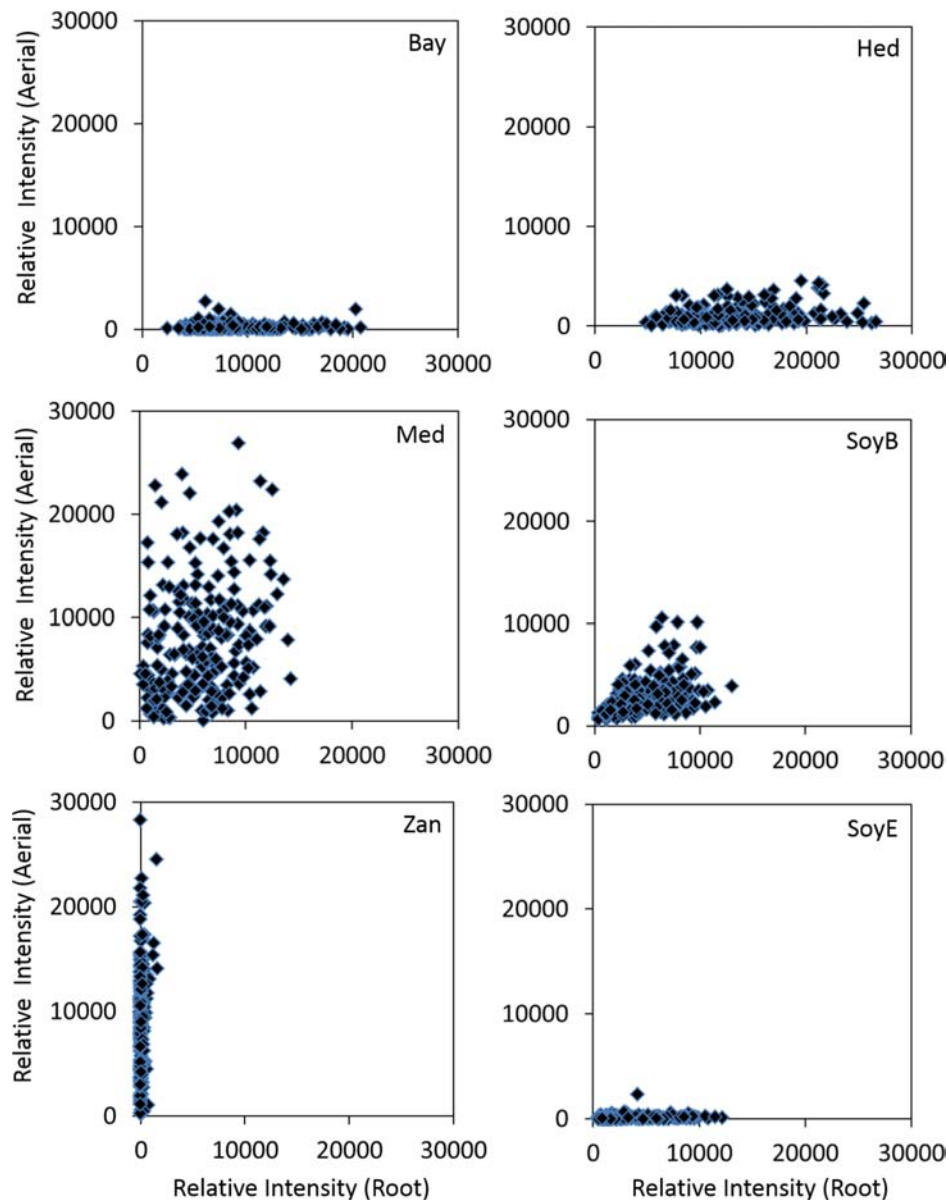


FIGURE 5 | Accumulations of various classes of saponins in the aerial (Y-axis) and root (X-axis) tissues of different *M. truncatula* ecotypes. Bay, bayogenin glycosides; Hed, hederagenin glycosides; Med, medicagenic acid glycosides; SoyB, soyasaponin B glycosides; Zan, zanhic acid glycosides; SoyE, soyasaponin E glycosides. Zanhic acid glycosides were predominantly found in the aerial tissues while bayogenin and hederagenin glycosides were mostly accumulated in the roots suggesting differential physiological roles for these compounds. In contrast, medicagenic acid glycosides are found in both the aerial and root tissues.

little effect on larvae development although they inhibited their growth (Nozzolillo et al., 1997). Feeding *Spodoptera littoralis* (Egyptian Cotton Leafworm) larvae with a diet supplemented with saponins isolated from alfalfa also significantly reduced their growth, fecundity and fertility and increased their mortality (Adel et al., 2000). Medicagenic acid and its glycosides were found to be much more effective in inhibiting larvae than hederagenin and its glycosides which are normally accumulated in roots. In a more recent study, the number of aphids infesting alfalfa was found to be inversely related to the contents of zanhic acid and medicagenic acid glycosides in the leaves (Goławska et al., 2012).

It was also demonstrated that these compounds were induced upon aphid infestation, indicating their anti-feeding properties. All these suggest that saponins in leaves have been tailored to defend against herbivore feeding through increased bitterness and toxicity.

Compared to leaf saponins, root saponins consisted of a different set of triterpene glycosides, with the major difference being the absence of zanhic acid glycosides and the presence of bayogenin glycosides and hederagenin glycosides (Figure 5). This suggests that P450s responsible for converting medicagenic acid to zanhic acid is not active in roots. This is not surprising as

leaves and roots are in different environments and may require different saponins to respond to different abiotic and biotic stress. The absence of the bitter zanhic acid glycosides in roots suggests that anti-feeding is less important in roots than in leaves. The abundance of bayogenin and hederagenin glycosides found in roots suggests that these root-specific compounds are important to root fitness. We hypothesize that they likely serve important roles in defense against soil pathogenic microbes. Indeed, saponins from *Medicago hybrida* roots were found to substantially inhibit six pathogenic fungi *Botrytis cinerea*, *Botrytis tulipae*, *Fusarium oxysporum* f. sp. *callistephi*, *F. oxysporum* f. sp. *narcissi*, *Phoma narcissi*, and *F. oxysporum* Schlecht (Saniewska et al., 2006). Similarly, saponins isolated from roots of alfalfa were also found to have strong anti-fungal activity (Jarecka et al., 2008). The strong antifungal activity was attributed to some hederagenin and medicagenic acid glycosides (Saniewska et al., 2006; Jarecka et al., 2008). Bayogenin glycosides have also been reported to possess anti-fungal activity (Martyniuk and Biały, 2008). Their inhibitory effects against *Cephalosporium gramineum*, a soil fungus that infects roots of plants, were markedly higher than their similar hederagenin glycoside counterparts (Martyniuk and Biały, 2008). Bayogenin differs with hederagenin only in that it possesses a hydroxyl group at C2 position (Figure 1). The hydroxyl groups at C2 and C3 positions are important for antifungal activities, possibly due to the increased polarity and solubility (Oleszek, 1996). Selective methylation of the hydroxyl groups in medicagenic acid showed that the hydroxyl group at C3 position is essential for antifungal activity (Levy et al., 1989). While glycosylation of the C3 hydroxyl group did not affect the overall antifungal activity, methylation or acetylation resulted in a significant loss of the antifungal activity. This suggests that the polarity of these compounds is important in their antifungal activity. The spatially differential accumulation of saponins in the aerial parts and roots shows that plant secondary metabolism is flexible and adaptable. Different tissues of the same plant can accumulate different sets of metabolites to increase their fitness and adapt to the environments. In *M. truncatula*, the bitter herbivore deterrent zanhic acid glycosides are predominantly found in the aerial tissues while the anti-fungal agents such as hederagenin and bayogenin glycosides are mostly found in roots to defend against soil borne fungi. In contrast, medicagenic acid glycosides, the broad spectrum and strong anti-microbial compounds, are accumulated in both the aerial and root tissues.

CONCLUSION

Saponin profiling of 201 *M. truncatula* ecotypes revealed a clear differential spatial accumulation of saponins in the aerial parts relative to the roots. Zanhic acid glycosides were only found in the aerial tissues. In contrast, bayogenin glycosides, hederagenin glycosides, and soyasaponin B glycosides were predominantly accumulated in roots, thus suggesting interesting ecological roles for these compounds in plant defense. Overall, zanhic acid glycosides, medicagenic acid glycosides and soyasaponin B glycosides were the three major triterpene saponins found in the aerial parts, while medicagenic acid glycosides, bayogenin

glycosides, hederagenin glycosides, SoyE glycosides, and SoyB glycosides constituted the major root saponins. Although medicagenic acid glycosides were found in both the aerial and root tissues, aerial parts were found to contain more medicagenic acid glycosides. Significant correlation between the quantity of saponins and the ecotypes' country of origin was observed. The European ecotypes were found to contain higher content of saponins than most of the African and the Middle-East ecotypes. This dataset also represents an extremely valuable resource for discovery of biosynthetic genes and deciphering saponin biosynthesis. Based on the correlation analysis of the metabolic profiling data and gene expression data, a number of P450 genes have been selected for further characterization to elucidate their roles in saponin biosynthesis in *M. truncatula* (Tzin et al., 2019).

DATA AVAILABILITY

All processed data for this study are included in the manuscript and/or the **Supplementary Files**.

AUTHOR CONTRIBUTIONS

LS conceptualized and designed the experiments, and secured the funding. BW and DY grew, harvested, dried, and ground the samples. DH extracted the ground samples and collected the LCMS data. ZL processed the data and drafted the manuscript. All authors were involved in the implementation of the experimental design and contributed to the final version of the manuscript.

FUNDING

The Sumner laboratory has been graciously supported by several entities over the years for the development of natural products profiling and plant metabolomics. For this project, these specifically included the NSF MCB Award 1024976, the Oklahoma Center for the Advancement of Science and Technology (OCAST) #PSB10-027. The Sumner lab was also supported by the University of Missouri, The Samuel Roberts Noble Foundation, the Bruker Daltonics GmbH, the NSF-JST Metabolomics for a Low Carbon Society, IOS Award IOS-1139489 and IOS-1639618, the NSF MRI DBI Award 1126719, and the NSF RCN Award 1340058.

ACKNOWLEDGMENTS

We gratefully acknowledge Shelagh Henson for her assistance in the plant care and harvest.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.00850/full#supplementary-material>

REFERENCES

- Adel, M. M., Sehnal, F., and Jurzysta, M. (2000). Effects of alfalfa saponins on the moth *Spodoptera littoralis*. *J. Chem. Ecol.* 26, 1065–1078.
- Agrall, J., Oleszek, W., Stochmal, A., Olsen, M., and Anderson, P. (2003). Herbivore-induced responses in alfalfa (*Medicago sativa*). *J. Chem. Ecol.* 29, 303–320.
- Avato, P., Bucci, R., Tava, A., Vitali, C., Rosato, A., Bialy, Z., et al. (2006). Antimicrobial activity of saponins from *Medicago* sp.: structure-activity relationship. *Phytother. Res.* 20, 454–457.
- Ben, C., Toueni, M., Montanari, S., Tardin, M. C., Fervel, M., Negahi, A., et al. (2013). Natural diversity in the model legume *Medicago truncatula* allows identifying distinct genetic mechanisms conferring partial resistance to Verticillium wilt. *J. Exp. Bot.* 64, 317–332. doi: 10.1093/jxb/ers337
- Bialy, Z., Jurzysta, M., Oleszek, W., Piacente, S., and Pizza, C. (1999). Saponins in alfalfa (*Medicago sativa* L.) root and their structural elucidation. *J. Agric. Food Chem.* 47, 3185–3192.
- Biazzi, E., Carelli, M., Tava, A., Abbruscato, P., Losini, I., Avato, P., et al. (2015). CYP72A67 catalyzes a key oxidative step in *Medicago truncatula* hemolytic saponin biosynthesis. *Mol. Plant* 8, 1493–1506. doi: 10.1016/j.molp.2015.06.003
- Carelli, M., Biazzi, E., Panara, F., Tava, A., Scaramelli, L., Porceddu, A., et al. (2011). *Medicago truncatula* CYP716A12 is a multifunctional oxidase involved in the biosynthesis of hemolytic saponins. *Plant cell* 23, 3070–3081. doi: 10.1105/tpc.111.087312
- Confalonieri, M., Cammareri, M., Biazzi, E., Pecchia, P., Salema Feveireiro, M. P., Balestrazzi, A., et al. (2009). Enhanced triterpene saponin biosynthesis and root nodulation in transgenic barrel medic (*Medicago truncatula* Gaertn.) expressing a novel β -amyrin synthase (AsOXA1) gene. *Plant Biotechnol. J.* 7, 172–182. doi: 10.1111/j.1467-7652.2008.00385.x
- Da Silva, P., Eyraud, V., Carre-Pierrat, M., Sivignon, C., Rahioui, I., Royer, C., et al. (2012). High toxicity and specificity of the saponin 3-GlcA-28-ArRhxyl-medicagenate, from *Medicago truncatula* seeds, for *Sitophilus oryzae*. *BMC Chem. Biol.* 12:3. doi: 10.1186/1472-6769-12-3
- Dixon, R. A., and Sumner, L. W. (2003). Legume natural products: understanding and manipulating complex pathways for human and animal health. *Plant Physiol.* 131, 878–885.
- Food and Drug Administration [FDA] (2001). *FDA Guidance for Industry: Bioanalytical Method Validation*. Silver Spring, MA: FDA.
- Francis, G., Kerem, Z., Makkar, H. P., and Becker, K. (2002). The biological action of saponins in animal systems: a review. *Br. J. Nutr.* 88, 587–605.
- Fukushima, E. O., Seki, H., Sawai, S., Suzuki, M., Ohyama, K., Saito, K., et al. (2013). Combinatorial biosynthesis of legume natural and rare triterpenoids in engineered yeast. *Plant Cell Physiol.* 54, 740–749. doi: 10.1093/pcp/pct015
- Golawska, S., Łukasik, I., Wójcicka, A., and Sytykiewicz, H. (2012). Relationship between saponin content in alfalfa and aphid development. *Acta Biol. Crac. Ser. Bot.* 54, 39–46. doi: 10.2478/v10182-012-0022-y
- Hoffmann, B., Trinh, T. H., Leung, J., Kondorosi, A., and Kondorosi, E. (1997). A new *medicago truncatula* line with superior in vitro regeneration, transformation, and symbiotic properties isolated through cell culture selection. *Mol. Plant Microbe Interact.* 10, 307–315.
- Huhman, D. V., Berhow, M. A., and Sumner, L. W. (2005). Quantification of saponins in aerial and subterranean tissues of *Medicago truncatula*. *J. Agric. Food Chem.* 53, 1914–1920.
- Jarecka, A., Saniewska, A., Biały, Z., and Jurzysta, M. (2008). The effect of *Medicago arabica*, *M. hybrida* and *M. sativa* saponins on the growth and development of *Fusarium oxysporum* schlecht f. sp. tulipae Apt. *Acta Agrobot.* 61:9.
- Kapusta, I., Janda, B., Stochmal, A., and Oleszek, W. (2005a). Determination of saponins in aerial parts of barrel medic (*Medicago truncatula*) by liquid chromatography-electrospray ionization/mass spectrometry. *J. Agric. Food Chem.* 53, 7654–7660.
- Kapusta, I., Stochmal, A., Perrone, A., Piacente, S., Pizza, C., and Oleszek, W. (2005b). Triterpene saponins from barrel medic (*Medicago truncatula*) aerial parts. *J. Agric. Food Chem.* 53, 2164–2170.
- Kirwan, J. A., Broadhurst, D. I., Davidson, R. L., and Viant, M. R. (2013). Characterising and correcting batch variation in an automated direct infusion mass spectrometry (DIMS) metabolomics workflow. *Anal. Bioanal. Chem.* 405, 5147–5157. doi: 10.1007/s00216-013-6856-7
- Levy, M., Zehavi, U., Naim, M., Polacheck, I., and Evron, R. (1989). Structure-biological activity relationships in alfalfa antimycotic saponins: the relative activity of medicagenic acid and synthetic derivatives thereof against plant pathogenic fungi. *J. Phytopathol.* 125, 209–216.
- Lima, F. G., Haraguchi, M., Pfister, J. A., Guimaraes, V. Y., Andrader, D. D. F., Ribeiro, C. S., et al. (2012). Weather and plant age affect the levels of steroidal saponin and *Pithomyces chartarum* spores in *Brachiaria* grass. *Int. J. Poisonous Plant Res.* 2:8.
- Martyniuk, S., and Bialy, Z. (2008). Effects of saponins from *Medicago arabica* on in vitro growth of *Cephalosporium gramineum*. *Phytopathol. Pol.* 49, 49–55.
- Moses, T., Papadopoulou, K. K., and Osbourn, A. (2014). Metabolic and functional diversity of saponins, biosynthetic intermediates and semi-synthetic derivatives. *Crit. Rev. Biochem. Mol. Biol.* 49, 439–462. doi: 10.3109/10409238.2014.953628
- Nozzolillo, C., Arnason, J. T., Campos, F., Donskov, N., and Jurzysta, M. (1997). Alfalfa leaf saponins and insect resistance. *J. Chem. Ecol.* 23, 995–1002.
- Oleszek, W. (1996). “Alfalfa saponins: structure, biological activity, and chemotaxonomy,” in *Saponins Used in Food and Agriculture*, eds G. R. Waller and K. Yamasaki (New York, NY: Plenum Press).
- Oleszek, W., Jurzysta, M., Ploszynski, M., Colquhoun, I. J., Price, K. R., and Fenwick, K. R. (1992). Zanic acid tridesmoside and other dominant saponins from alfalfa (*Medicago sativa* L.) aerial parts. *J. Agric. Food Chem.* 40, 191–196.
- Oleszek, W., Nowacka, J., Gee, J. M., Wortley, G. M., and Johnson, I. T. (1994). Effects of some purified alfalfa (*Medicago sativa*) saponins on transmembrane potential difference in mammalian small intestine. *J. Sci. Food Agric.* 65, 35–39.
- Oleszek, W., Price, K. R., Colquhoun, I. J., Jurzysta, M., Ploszynski, M., and Fenwick, G. R. (1990). Isolation and identification of alfalfa (*Medicago sativa* L.) root saponins: their activity in relation to a fungal bioassay. *J. Agric. Food Chem.* 38, 1810–1817.
- Paape, T., Bataillon, T., Zhou, P., Kono, T. J. Y., Briskine, R., Young, N. D., et al. (2013). Selection, genome-wide fitness effects and evolutionary rates in the model legume *Medicago truncatula*. *Mol. Ecol.* 22, 3525–3538. doi: 10.1111/mec.12329
- Perez, A. J., Kowalczyk, M., Simonet, A. M., Macias, F. A., Oleszek, W., and Stochmal, A. (2013). Isolation and structural determination of triterpenoid glycosides from the aerial parts of alsike clover (*Trifolium hybridum* L.). *J. Agric. Food Chem.* 61, 2631–2637. doi: 10.1021/jf305541e
- Pollier, J., Morreel, K., Geelen, D., and Goossens, A. (2011). Metabolite profiling of triterpene saponins in *Medicago truncatula* hairy roots by liquid chromatography Fourier transform ion cyclotron resonance mass spectrometry. *J. Nat. Prod.* 74, 1462–1476. doi: 10.1021/np200218r
- Ronfort, J., Bataillon, T., Santoni, S., Delalande, M., David, J., and Prosperi, J. M. (2006). Microsatellite diversity and broad scale geographic structure in a model legume: building a set of nested core collection for studying naturally occurring variation in *Medicago truncatula*. *BMC Plant Biol.* 6:28. doi: 10.1186/1471-2229-6-28
- Saniewska, A., Jarecka, A., Biały, Z., and Jurzysta, M. (2006). Antifungal activity of saponins originated from *Medicago hybrida* against some ornamental plant pathogens. *Acta Agrobot.* 59, 51–58.
- Sen, S., Makkar, H. P., and Becker, K. S. (1998). Alfalfa saponins and their implication in animal nutrition. *J. Agric. Food Chem.* 46, 131–140.
- Soliz-Guerrero, J. B., Rodríguez-García, D., Rodríguez-García, R., Angulo-Sánchez, L., and Méndez-Padilla, G. (2002). *Quinoa Saponins: Concentration and Composition Analysis*. Alexandria, VA: ASHS Press.
- Suzuki, H., Achnine, L., Xu, R., Matsuda, S. P., and Dixon, R. A. (2002). A genomics approach to the early stages of triterpene saponin biosynthesis in *Medicago truncatula*. *Plant J.* 32, 1033–1048.
- Szakiel, A., Pączkowski, C., and Henry, M. (2011). Influence of environmental abiotic factors on the content of saponins in plants. *Phytochem. Rev.* 10, 471–491.
- Tava, A., Mella, M., Avato, P., Argentieri, M. P., Bialy, Z., and Jurzysta, M. (2005). Triterpenoid glycosides from leaves of *Medicago arborea* L. *J. Agric. Food Chem.* 53, 9954–9965.
- Tava, A., and Pecetti, L. (2012). Chemical investigation of saponins from twelve annual *Medicago* species and their bioassay with the brine shrimp *Artemia salina*. *Nat. Prod. Commun.* 7, 837–840.
- Tava, A., Scotti, C., and Avato, P. (2011). Biosynthesis of saponins in the genus *Medicago*. *Phytochem. Rev.* 10, 459–469.

- Tzin, V., Snyder, J. H., Yang, D. S., Huhman, D. V., Watson, B. S., Allen, S. N., et al. (2019). Integrated metabolomics Identifies CYP72A67 and CYP72A68 Oxidases in the biosynthesis of medicago truncatula oleanate sapogenins. *Metabolomics* 15:85. doi: 10.1007/s11306-019-1542-1
- Voutquenne, L., Lavaud, C., Massiot, G., and Men-Olivier, L. L. (2002). Structure-activity relationships of haemolytic saponins. *Pharm. Biol.* 40, 253–262.
- Zelena, E., Dunn, W. B., Broadhurst, D., Francis-McIntyre, S., Carroll, K. M., Begley, P., et al. (2009). Development of a robust and repeatable UPLC-MS method for the long-term metabolomic study of human serum. *Anal. Chem.* 81, 1357–1364. doi: 10.1021/ac8019366
- Zhu, Z., Liang, Z., Han, R., and Wang, X. (2009). Impact of fertilization on drought response in the medicinal herb *Bupleurum chinense* DC.: growth and saikosaponin production. *Ind. Crops Prod.* 29, 629–633.
- Zimmer, D. E., Pedersen, M. W., and McGuire, C. F. (1967). A bioassay for *Alfalfa Saponins* using the fungus, *Trichoderma viride* pers. ex Fr.1. *Crop Sci.* 7, 223–224.
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Copyright © 2019 Lei, Watson, Huhman, Yang and Sumner. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Assessing Chemical Diversity in *Psilotum nudum* (L.) Beauv., a Pantropical Whisk Fern That Has Lost Many of Its Fern-Like Characters

Dunja Šamec^{1,2}, Verena Pierz^{1,3}, Narayanan Srividya¹, Matthias Wüst³ and B. Markus Lange^{1*}

¹ Institute of Biological Chemistry and M.J. Murdock Metabolomics Laboratory, Washington State University, Pullman, WA, United States, ² Ruder Bošković Institute, Zagreb, Croatia, ³ Chair of Bioanalytics, Institute of Nutritional and Food Sciences, University of Bonn, Bonn, Germany

OPEN ACCESS

Edited by:

Kazuki Saito,
RIKEN Center for Sustainable
Resource Science (CSRS), Japan

Reviewed by:

A. Daniel Jones,
Michigan State University,
United States
Zhigang Yang,
Lanzhou University, China

*Correspondence:

B. Markus Lange
lange-m@wsu.edu

Specialty section:

This article was submitted to
Plant Metabolism
and Chemodiversity,
a section of the journal
Frontiers in Plant Science

Received: 12 March 2019

Accepted: 18 June 2019

Published: 09 July 2019

Citation:

Šamec D, Pierz V, Srividya N,
Wüst M and Lange BM (2019)
Assessing Chemical Diversity
in *Psilotum nudum* (L.) Beauv.,
a Pantropical Whisk Fern That Has
Lost Many of Its Fern-Like
Characters. *Front. Plant Sci.* 10:868.
doi: 10.3389/fpls.2019.00868

Members of the Psilotales (whisk ferns) have a unique anatomy, with conducting tissues but lacking true leaves and roots. Based on recent phylogenies, these features appear to represent a reduction from a more typical modern fern plant rather than the persistence of ancestral features. In this study, extracts of several *Psilotum* organs and tissues were analyzed by Gas Chromatography – Mass Spectrometry (GC-MS) and High Performance Liquid Chromatography – Quadrupole Time of Flight – Mass Spectrometry (HPLC-QTOF-MS). Some arylpyrones and biflavonoids had previously been reported to occur in *Psilotum* and these metabolite classes were found to be prominent constituents in the present study. Some of these were enriched and further characterized by Nuclear Magnetic Resonance (NMR) spectroscopy. HPLC-QTOF-MS and NMR data were searched against an updated Spektraris database (expanded by incorporating over 300 new arylpyrone and biflavonoid spectral records) to aid significantly with peak annotation. Principal Component Analysis (PCA) with combined GC-MS and HPLC-QTOF-MS data sets obtained with several *Psilotum* organs and tissues indicated a clear separation of the sample types. The principal component scores for below-ground rhizome samples corresponded to the vectors for carbohydrate monomers and dimers and small organic acids. Above-ground rhizome samples had principal component scores closer to the direction of vectors for arylpyrone glycosides and sucrose (which had high concentrations in above- and below-ground rhizomes). The unique position of brown synangia in a PCA plot correlated with the vector for biflavonoid glycosides. Principal component scores for green and yellow synangia correlated with the direction of vectors for arylpyrone glycosides and biflavonoid aglycones. Localization studies with cross sections of above-ground rhizomes, using Matrix-Assisted Laser Desorption/Ionization – Mass Spectrometry (MALDI-MS), provided evidence for a preferential accumulation of arylpyrone glycosides and biflavonoid aglycones in cells

of the chlorenchyma. Our results indicate a differential localization of metabolites with potentially tissue-specific functions in defenses against biotic and abiotic stresses. The data are also a foundation for follow-up work to better understand chemical diversity in the Psilotales and other members of the fern lineage.

Keywords: arylpyrone, biflavonoid, mass spectrometry, metabolomics, nuclear magnetic resonance, whisk fern

INTRODUCTION

Free-sporing vascular plants encompass two distinct evolutionary lineages, the lycophytes and ferns, with the latter resolved as more closely related to seed plants (Kenrick and Crane, 1997; Pryer et al., 2001). Whisk ferns (order Psilotales), which comprise two genera (*Psilotum* and *Tmesipteris*) in the family Psilotaceae, have conducting tissues but no veins, and lack true leaves and roots. Water and mineral absorption occurs through underground, horizontally creeping rhizomes, sometimes in association with symbiotic fungi (mycorrhizae) (Ducket and Ligrone, 2005). Plants grow mostly as epiphytes (using other plants as physical support) in moist habitats. The stem-like aerial portion of rhizomes of members of the Psilotaceae is covered by an epidermis, followed inward by extensive cortical areas, a single-layered endodermis, and a thick-walled protostele that accommodates the water and nutrient-conducting tissues (Pittermann et al., 2011). The epidermal layer of the photosynthetic above-ground rhizomes contains stomata for gas exchange (Nilsen, 1995). In the genus *Psilotum*, above-ground rhizomes have many branches with scale-like appendages called enations. These structural outgrowths resemble miniature leaves but, unlike true leaves, have no internal vascular tissues. Above these enations, positioned laterally along the distal portions of aerial shoots, are spore-containing syngangia, which result from the fusion of three adjacent sporangia (Renzaglia et al., 2001).

Because of its unusual anatomical characteristics, *P. nudum* was traditionally thought to be descended from the earliest vascular plants (Banks, 1975), and conflicting views regarding the placement of the Psilotales remained in the literature for decades. Recent phylogenies based on both morphological characters and extensive sequence data provided strong evidence that Psilotales, Ophioglossales (moonworts) and Marattiales (king ferns) – all eusporangiate ferns – form a monophyletic clade that is sister to leptosporangiate ferns, the largest group of living ferns (Doyle, 2018; Rothwell et al., 2018). The unique anatomy of extant Psilotales therefore appears to represent a reduction from a more typical modern fern plant rather than the persistence of ancestral features. While recent progress has been made with regard to resolving the classification of vascular plants, there is still a notable lack of knowledge regarding the phytochemical diversification associated with the adaptive radiation of ferns.

Abbreviations: AMT tag, accurate mass-time tag; CHCA, α -cyano-4-hydroxycinnamic acid; DBA, 2,5-dihydroxybenzoic acid; GC-MS, gas chromatography – mass spectrometry; HPLC-QTOF-MS, high performance liquid chromatography – quadrupole time of flight – mass spectrometry; MALDI-MS, matrix-assisted laser desorption/ionization – mass spectrometry; NMR, nuclear magnetic resonance; PC, principal component; PCA, principal component analysis.

We selected *Psilotum nudum* (L.) Beauv. to evaluate chemical diversity in the fern lineage, as only limited knowledge exists on this topic. Psilotin and 3'-hydroxypsilotin are unusual C₁₁ arylpyrone glycosides unique to the Psilotaceae (McInnes et al., 1965; Tse and Towers, 1967; Balza et al., 1985; Takahashi et al., 1990). Psilotic acid is a C₆-C₄ organic acid that is structurally related to the psilotin aglycone (psilotinin) (Shamsuddin et al., 1985). Prominent flavonoid glycosides in the Psilotaceae are O-glucosides of the biflavonoid, amentoflavone, and C- and O-glycosides of the flavone, apigenin (Cooper-Driver, 1977; Wallace and Markham, 1978; Markham, 1984). A survey across sixteen pteridophytes (ferns and fern allies), including *P. nudum*, concluded that the sterol composition is generally similar to that of spermatophytes (seed plants), with β -sitosterol, campesterol and stigmaterol as principal constituents (Chiu et al., 1988). *P. nudum* tissues were also demonstrated to contain representatives of several phytohormone classes (auxins, cytokinins and gibberellins) (Takahashi et al., 1984; Abul et al., 2010). In this pilot study, which is the beginning of efforts to chart out the most abundant classes of specialized metabolites in ferns, we demonstrate the utility of multi-platform analyses for capturing the unique chemical fingerprints of different *P. nudum* organs and tissues. In addition, we report the tissue-level localization of the most prominent arylpyrone glycoside and biflavonoid constituents.

MATERIALS AND METHODS

Chemicals and Solvents

Solvents for extraction and chromatography were of the highest commercial grade and obtained from Sigma-Aldrich (St. Louis, MO, United States). Deuterated solvents for nuclear magnetic resonance (NMR) spectroscopy were obtained from Cambridge Isotope Laboratories Inc. (Andover, MA, United States), with details in **Table 2**. All authentic standards, reference materials (red phosphorus, α -cyano-4-hydroxycinnamic acid, 9-anthracenecarboxylic acid, sinapic acid and vanillic acid) and reagents (*N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide) were generally purchased from Sigma-Aldrich (St. Louis, MO, United States); exceptions: 2,5-dihydroxybenzoic acid (TCI America, Portland, OR, United States) and leucine enkephalin (Waters, Milford, MA, United States).

Plant Growth

Psilotum nudum (L.) P. Beauv. plants had been established from rhizomes roughly 6 years before the initiation of the experiments described here. A voucher specimen was deposited with the John G. Searle Herbarium of the Field Museum (Chicago,

IL, United States). Plants were maintained in a greenhouse under ambient lighting, with supplemental lighting during winter months provided by high-intensity discharge lamps. The daily light integral varied from 15 to 25 mol m⁻² d⁻¹. Temperatures ranged between 22 and 27°C and the humidity was set to 70%. At the time of harvesting, *P. nudum* produced synangia that, based on color (green, yellow or brown), could be differentiated into three developmental stages (immature, mature, and senescent). Five biological replicates were harvested at the same time of day (11:00 AM, Pacific Daylight Savings Time) for the following organs: below-ground rhizome, above-ground rhizome (stem), and (separately) green, yellow and brown synangia. Samples were immediately frozen in liquid nitrogen, freeze-dried (aerial parts for 5 days, rhizomes for 7 days). Lyophilized material was homogenized to a fine powder under liquid nitrogen using mortar and pestle. Defined quantities of homogenate were weighed out, placed in a 2 ml microfuge tube, and stored as aliquots at -20°C until further use.

Metabolite Extraction and Derivatization for Analysis by Gas Chromatography – Mass Spectrometry

Frozen tissue homogenate from each sample (15 ± 3 mg) was transferred to 8 ml glass tubes and overlaid with 700 µL methanol (containing myristic acid-d₂₇ (CDN Isotopes, Quebec, Canada) as internal standard at 1.5 mg/ml) and 25 µL water. Tubes were capped tightly and heated in a water bath to 70°C for 15 min, centrifuged for 2 min at 3,500 × g, and supernatants transferred to new 8 ml glass vials. To each supernatant, 700 µL of water and 375 µL of chloroform were added and the contents of the tube mixed with a multi-tube vortexer (VWR Scientific, South Plainfield, NY, United States) for 15 min at a speed setting of 4. Extracts were centrifuged for 15 min at 3,500 × g, the upper aqueous phase was combined with the first methanol extract (henceforth referred to as aqueous methanol extract), and the lower organic phase was collected separately (chloroform extract). The two extracts were separately evaporated to dryness [Vacufuge Plus for aqueous methanol extract (Eppendorf, Hauppauge, NY, United States); EZ-Bio Evaporator for chloroform extract (GeneVac LTD, Ipswich, United Kingdom)]. Dried samples were derivatized just-in-time by adding 10 µL of a 40 mg/ml solution of methoxyamine hydrochloride in pyridine and shaking gently at 30°C for 90 min, then adding 50 µL of *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide (MSTFA; Sigma-Aldrich, St. Louis, MO, United States) and shaking gently at 37°C for 30 min. Samples were allowed to cool to room temperature, and the extract was transferred to a glass insert, which was then placed in a 2 ml glass reaction vial.

Gas Chromatography – Mass Spectrometry Analysis

Gas chromatography – mass spectrometry (GC–MS) was performed under the following conditions: injection volume: 1 µL (splitless mode); GC instrument: 6890N (Agilent

Technologies, Santa Clara, CA, United States), GC; column: DB-5MS + DG (30 m × 0.25 mm × 0.25 µm; J&W Scientific, Santa Clara, CA, United States); inlet temperature: 250°C; temperature program: start at 60°C, ramp to 320°C at 3°C/min, hold for 10 min; retention time locking: myristic acid-d₂₇ at 42.06 min at an inlet pressure of 10.65 psi; MS instrument: 5975 MSD (Agilent Technologies, Santa Clara, CA, United States); transfer line temperature: 250°C; electron ionization at 70 eV. Data analysis was performed using ChemStation, version E.02.00.493 (Agilent Technologies, Santa Clara, CA, United States). Custom spectral databases (specifying retention time, a quantification signal and three qualifier ions) were created using authentic standards from our in-house library for the identification of GC–MS peaks (Supplementary Table S1). Peaks generated by unidentified analytes were annotated based on community reporting guidelines (Bino et al., 2004; Fiehn et al., 2007). Raw data values were normalized for sample weight and signal intensity associated with the internal standard. Normalized data values were *z*-transformed (autoscaled) prior to statistical analyses.

Metabolite Extraction for High Performance Liquid Chromatography – Quadrupole Time-of-Flight – Mass Spectrometry

Frozen tissue homogenate from each sample (30 ± 5 mg) was transferred to a 2 ml reaction tube and extracted with 1 ml of 80% aqueous methanol (containing 6 mg/L anthracene-9-carboxylic acid as internal standard) for 10 min [multi-tube vortexer (VWR Scientific, South Plainfield, NY, United States) at highest speed setting] and subsequent sonication for 20 min (ultrasonic bath at highest intensity setting, Fisher Scientific, Hampton, NY, United States). Following centrifugation for 10 min at 13,000 × g, supernatants were filtered through 0.22 µm polypropylene filter material and collected in plastic inserts for 2 ml reaction vials.

High Performance Liquid Chromatography – Quadrupole Time-of-Flight – Mass Spectrometry Analysis

High Performance Liquid Chromatography – Quadrupole Time-of-Flight – Mass Spectrometry (HPLC–QTOF–MS) was performed under the following conditions: HPLC system: 1290 system (Agilent Technologies, Santa Clara, CA, United States) consisting of thermo-controlled autosampler (set to 4°C), binary pump (operated at 0.6 ml/min), isocratic pump [operated at 0.1 ml/min flow rate to introduce a reference mass solution containing 300 nM purine (exact mass 120.043596 g/mol) and 250 nM hexakis-(1H, 1H,3H-tetrafluoropropoxy)-phosphazine (exact mass 921.002522 g/mol) in acetonitrile/water (95:5; v/v)], thermo-controlled column compartment (set to 60°C), and diode array detector (scanning range 210–600 nm, resolution 1.2 nm); injection volume: 10 µL; Column: HD Zorbax SB-Aq (100 × 2.1 × mm; 1.8 µm pore size, Agilent Technologies, Santa Clara, CA, United States); Mobile phase: 0.1% (v/v) formic acid in water (solvent A) and 0.1% (v/v) formic acid

in acetonitrile (solvent B). Gradient: 5% B at start; linear gradients to 10% B at 5 min, 20% B at 10 min, 80% B at 35 min, 95% B at 45 min; QTOF-MS instrument: 6530 series with electrospray ion source (Agilent Technologies, Santa Clara, CA, United States); polarity: positive; drying gas flow rate: 10 L/min; drying gas temperature: 325°C; nebulizer pressure: 2.4 bar; m/z range: 100–1,200 (high gain mode); scan rate: 1.4 scans/s for MS and 4 scans/s for MS/MS. Data analysis was performed using the MassHunter Workstation software package [B.07.00, Qualitative Analysis and B.06.00, Profinder, Agilent Technologies, Santa Clara, CA, United States). For each detected peak, molecular feature extraction (considering retention time (tolerance window 1.30 s) and high mass accuracy (m/z tolerance window 10 ppm)], deconvolution, and alignment across samples were performed using the recursive feature extraction algorithm (settings: threshold of 10,000 counts and peak spacing tolerance of 0.0025 m/z). Quasi-molecular ions and adducts were considered ($[M+H]^+$, $[M+Na]^+$, $[M+K]^+$, $[M+NH_4]^+$), as were the corresponding dimers. The minimum absolute height required for feature extraction in the recursive step was set to 10,000 counts (sum of all peaks for a given molecular entity), which had to be fulfilled in at least three of five biological replicates. The global filter was limited to 2,000 results. Peak annotation was performed based on a combination of chromatographic, mass spectral (accurate mass and MS/MS fragmentation patterns), evaluation of the literature, and searches against spectral databases (Table 1). Peaks generated by unidentified analytes were annotated based on community reporting guidelines (Bino et al., 2004; Fiehn et al., 2007). MS/MS spectra for identified peaks were submitted to MassBank (Horai et al., 2010) to expand a widely used community spectral resource. Normalized data values for HPLC-QTOF-MS peaks were z-transformed (autoscaled) and combined with the normalized and z-transformed GC-MS data (Supplementary Table S2). The combined HPLC-QTOF-MS and GC-MS data set were processed by Principal Component Analysis (PCA) using the R statistical package¹, for which the settings and outcomes are summarized in Supplementary Table S3.

Metabolite Isolation and Analysis by Nuclear Magnetic Resonance Spectroscopy

Above-ground biomass from *P. nudum* was harvested and homogenized to a fine powder in the presence of liquid nitrogen. A 300 mg aliquot of the homogenate was extracted with 10 ml of 80% aqueous methanol by vigorous mixing for 10 min (Vortex Mixer, VWR Scientific, South Plainfield, NY, United States; operated at highest speed setting) and subsequent sonication in an ultrasonic bath for 20 min. Following centrifugation of this mixture for 10 min at 13,000 rpm, the supernatant was recovered and filtered through a 0.22 µm polypropylene filter. The extract was stored at –20°C until further processing. Aliquots (100 µl each) of the filtered extracts were injected onto a C18 reversed phase and absorbance at 280 and 360 nm was monitored

(1100 Series HPLC system; Agilent Technologies, Santa Clara, CA, United States). The mobile phase consisted of two solvents (A: 0.2% (v/v) acetic acid in water; B: 0.2% (v/v) acetic acid in methanol) and the separation of metabolites was achieved using the following gradient: 2% B at start, with a series of linear gradients to 35% B at 10 min, 60% B at 21 min, 90% B at 40 min, and 98% B at 50 min. The flow rate was set to 1.3 ml/min. Trial runs indicated when metabolites of interest eluted and fractions were collected accordingly. The eluents of several runs were accumulated and each of these fractions evaporated to dryness in a rotary evaporator. Each residue was dissolved in a deuterated solvent and NMR spectra were acquired with the settings listed in Supplementary Table S4. Spectral records for bioflavonoids and arylpyrones were generated based on information extracted from the literature (listing in Supplementary Table S5) and integrated into the Spektraris database (Cuthbertson et al., 2013; Fishedick et al., 2015). The combined spectral data from HPLC-QTOF-MS and NMR were then used to search for matches in the Spektraris online resource (Table 2).

Metabolite Imaging by Matrix-Assisted Laser Desorption/Ionization – Mass Spectrometry

Psilotum nudum above-ground rhizomes were cross-sectioned into 2 cm segments, embedded in 3% (w/v) agarose, and stored at –80°C until further processing. On the day of the metabolite imaging analysis, the chamber of a CM 1950 Cryostat (Leica Biosystems, Buffalo Grove, IL, United States) was set to –20°C, embedded samples were sectioned to 30 µm thickness and sections immediately transferred to an imaging target plate (Waters Corp., Milford, MA, United States). The ionization matrices tested for their suitability with *P. nudum* metabolites were 2,5-dihydroxybenzoic acid (DBA), α-cyano-4-hydroxycinnamic acid (CHCA), sinapic acid and vanillic acid (each at 40 mg/ml (w/v) in methanol/water (1:1; v/v)). Matrices were applied with a sample preparation system (TM-Sprayer of HTX Technologies, Chapel Hill, NC, United States) connected to an 1100 Series HPLC Binary Pump (Agilent Technologies, Santa Clara, CA, United States). The settings were: flow rate at 0.05 ml/min; nozzle temperature at 80°C; spraying velocity at 1,250 mm/min; 12 passes; and track spacing of 1 mm. The final amount of matrix deposited per linear distance was 0.19 mg mm^{–2}. Besides matrix-covered samples, the following chemicals were also spotted onto the imaging target plates: red phosphorus for instrument calibration (10 mg/ml in acetone), leucine-enkephalin to generate a lock mass [10 mg/ml mixed with 3.4 mg/ml CHCA in methanol/water (1:1; v/v)], and authentic standards [1 mg/ml of amentoflavone, psilotin and 3'-hydroxypsilotin mixed with 5 mg/ml DHB in methanol/water (1:1; v/v)]. Metabolite imaging was performed by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry (MALDI-MS) on a Synapt G2-S instrument equipped with an ion mobility drift tube and operated with MassLynx software version 4.1 (Waters, Milford, MA, United States). The imaging target plate was introduced into the sample chamber and the laser operated with the following settings: 1,000 Hz firing rate; laser energy

¹<https://www.r-project.org/>

TABLE 1 | Annotation of HPLC-QTOF-MS peaks.

Accurate Mass – Time Tag	Monoisotopic Mass (Measured/Calculated)	Δ ppm	Molecular Formula	MS (ESI-Positive)	MS/MS(ESI-Positive) (Varying Collision Energies)	Annotation	References; Further Evidence
BML-LCMS18-4.66-368.1101	368.1101/368.1107	1.79	C17H20O9	[M+H] ⁺ 369.1198 [M+Na] ⁺ 391.0999 [M+K] ⁺ 407.0696 [2M+Na] ⁺ 759.2115 [2M+K] ⁺ 775.1595	10 eV: 391.0999 (100), 123.0436 (80), 207.0649 (55), 189.0564 (30) 50 eV: 123.0438 (100), 189.0540 (35), 227.0002 (34), 98.9747 (26), 199.0080 (23), 110.9751 (19), 115.0537 (18), 171.0427 (17), 147.0451 (16), 231.0246 (15)	3'-Hydroxypsilotin	Balza et al., 1985; NMR (Table 2; F2)
BML-LCMS18-4.76-530.1636	530.1636/530.1636	0.00	C23H30O14	[M+Na] ⁺ 553.1530	30 eV: 553.1504 (100), 207.1641 (11) 50 eV: 123.0433 (100), 173.0570 (48), 85.0281 (23), 189.0512 (29), 115.0534 (26), 203.0539 (17), 116.9908 (15), 147.0437 (14)	3'-Hydroxypsilotinin-di-O-hexoside	
BML-LCMS18-5.12-514.1688	514.1688/514.1686	0.08	C23H30O13	[M+H] ⁺ 515.1751 [M+Na] ⁺ 537.1591 [2M+Na] ⁺ 1051.3232	30 eV: 537.1603 (100), 191.0746 (40), 391.0731 (15), 173.0627 (14), 107.0520 (7) 50 eV: 107.0538 (100), 173.0650 (87), 201.0111 (26), 229.0230 (20), 117.0743 (12), 145.0677 (12), 98.9808 (11)	Psilotinin-di-O-hexoside I	
BML-LCMS18-5.20-352.1168	352.1168/352.1158	2.79	C17H20O8	[M+H] ⁺ 353.1250 [M+Na] ⁺ 375.1075 [2M+Na] ⁺ 727.2223	10 eV: 375.1075 (100), 173.0598 (70), 107.0487 (63), 191.0699 (45), 123.0434 50 eV: 107.0492 (100), 98.9753 (78), 173.0588 (66), 110.9858 (64), 183.0123 (44), 127.0544 (43), 167.0168 (42), 117.0706 (33), 201.0072 (31), 145.0635 (25), 229.0171 (23), 30eV: 207.0677 (100), 189.0546 (81), 123.0443 (62) 50 EV: 123.0451 (100), 189.0566 (29), 173.0595 (27), 147.0451 (12)	Psilotin	McInnes et al., 1965; Authentic Standard
BML-LCMS18-5.52-206.0568	206.0568/206.0580	0.12	C11H10O4	[M+H] ⁺ 207.0652 [M+Na] ⁺ 229.0475	30eV: 207.0677 (100), 189.0546 (81), 123.0443 (62) 50 EV: 123.0451 (100), 189.0566 (29), 173.0595 (27), 147.0451 (12)	3'-Hydroxypsilotinin	
BML-LCMS18-6.83-514.1686	514.1686/514.1686	0.00	C23H30O13	[M+H] ⁺ 515.1741 [M+Na] ⁺ 537.1575 [2M+Na] ⁺ 1051.3228	30 eV: 537.1757 (100) 50 eV: 173.0591 (100), 107.0476 (86), 201.0053 (69), 85.0275 (55), 229.0168 (49), 127.0540 (41), 145.0625 (36), 97.0267 (34)	Psilotinin-di-O-hexoside II	
BML-LCMS18-8.31-190.0627	190.0627/190.0631	0.20	C11H10O3	[M+H] ⁺ 191.0703 [M+Na] ⁺ 213.0528	30 eV: 173.0589 (100), 107.0491 (82) 50 eV: 107.0472 (100), 173.0592 (76), 145.0669 (18)	Psilotinin	
BML-LCMS18-11.20-594.1583	594.1583/594.1585	0.02	C27H30O15	[M+H] ⁺ 595.1654 [M+Na] ⁺ 617.1470	10 eV: no fragmentation 50 eV: 271.0583 (100)	Apigenin-6,8-di-C-glucoside (vicenin-2)	Markham, 1984; Authentic Standard

(Continued)

TABLE 1 | Continued

Accurate Mass – Time Tag	Monoisotopic Mass (Measured/Calculated)	Δppm	Molecular Formula	MS (ESI-Positive)	MS/MS(ESI-Positive) (Varying Collision Energies)	Annotation	References; Further Evidence
BML-LCMS18-13.99-1024.2462	1024.2462/1024.2485	2.21	C48H48O25	[M+H] ⁺ 1025.2536 [M+Na] ⁺ 1047.2336	10 eV: no fragmentation 50 eV: 539.0979 (100), 701.1499 (81), 269.1301(10)	Amentoflavone-tri-hexoside I	
BML-LCMS18-14.26-432.1067	432.1067/432.1056	2.44	C21H20O10	[M+H] ⁺ 433.1130 [M+Na] ⁺ 455.0945	10 eV: no fragmentation 50 eV: 271.0597 (100), 153.0175 (4)	Apigenin-7-O-glucoside (apigenitrin, cosmosin)	Wallace and Markham, 1978; Authentic Standard
BML-LCMS18-14.50-578.1636	578.1636/578.1636	0.00	C27H30O14	[M+H] ⁺ 579.1712 [M+Na] ⁺ 601.1520 [2M+Na] ⁺ 1179.3113	10 eV: no fragmentation 50 eV: 271.0598 (100)	Apigenin-7-O-rhamnoglucoside (rhofolin)	Wallace and Markham, 1978
BML-LCMS18-14.82-1024.2480	1024.2480/1024.2485	0.03	C48H48O25	[M+H] ⁺ 1025.2551 [M+Na] ⁺ 1047.2352	10 eV: no fragmentation 50 eV: 539.0966 (100), 701.1482 (73), 863.1989 (14)	Amentoflavone-tri-hexoside II	
BML-LCMS18-15.78-862.1946	862.1946/862.1956	1.21	C42H38O20	[M+H] ⁺ 863.2021 [M+Na] ⁺ 885.1831 [M+K] ⁺ 901.1516	10 eV: no fragmentation 50 eV: 539.0967 (100), 701.1482 (6)	Amentoflavone-di-hexoside I	
BML-LCMS18-16.60-862.1938	862.1938/862.1956	2.14	C42H38O20	[M+H] ⁺ 863.2010 [M+Na] ⁺ 885.1825	10 eV: no fragmentation 50 eV: 539.0963 (100), 701.1473 (34)	Amentoflavone-di-hexoside II	
BML-LCMS18-17.56-702.1582	702.1582/702.1585	0.38	C36H30O15	[M+H] ⁺ 703.1642 [M+Na] ⁺ 725.1487	10 eV: no fragmentation 50 eV: 421.0545 (100), 541.1130 (43), 311.0555 (17), 337.0343 (16), 271.0582 (10), 153.0175 (8), 137.0583 (7), 147.0434 (7), 297.0413 (6), 379.0452 (5)	Dihydroamentoflavone-hexoside I	
BML-LCMS18-18.12-700.1418	700.1418/700.1428	1.46	C36H28O15	[M+H] ⁺ 701.1490 [M+Na] ⁺ 723.1310	10 eV: no fragmentation 50 eV: 539.0976 (100), 403.0443 (12), 377.0654 (5)	Amentoflavone-hexoside I	
BML-LCMS18-18.28-702.1562	702.1562/702.1585	3.23	C36H30O15	[M+H] ⁺ 703.1632 [M+Na] ⁺ 725.1442	10 eV: no fragmentation 50 eV: 389.1015 (100), 541.1112 (65), 153.0958 (55), 415.0787 (20), 403.0439 (7)	Dihydroamentoflavone-hexoside II	
BML-LCMS18-19.16-700.1424	700.1424/700.1428	1.00	C36H28O15	[M+H] ⁺ 701.1493 [M+Na] ⁺ 723.1304	10 eV: no fragmentation 50 eV: 539.0965 (100), 403.0434 (19), 377.0658 (8), 153.017 (5)	Amentoflavone-hexoside II	
BML-LCMS18-19.96-554.0856	554.0856/554.0849	1.24	C30H18O11	[M+H] ⁺ 555.0931	10 eV: no fragmentation 50 eV: 403.0641 (100), 153.0641 (94), 405.0951 (78), 377.0672 (70), 121.0297 (51), 347.0569 (40), 375.0844 (39), 335.0551 (36), 271.0600 (25), 283.0604 (24), 555.0880 (24)	Hydroxy-amentoflavone	Zhang Y.X. et al., 2011
BML-LCMS18-20.73-540.1053	540.1053/540.1057	0.64	C30H20O10	[M+H] ⁺ 541.1125 [M+Na] ⁺ 563.0948 [2M+Na] ⁺ 1103.1994	10 eV: no fragmentation 50 eV: 311.0555 (100), 337.0350 (100), 283.0604 (94), 421.0558 (36), 312.0585 (19), 147.0433 (11), 335.055 (10), 253.0490 (8), 153.0189 (8),	2,3-Dihydro-amentoflavone	Zhang Y.X. et al., 2011 NMR (Table 2; F8)

(Continued)

TABLE 1 | Continued

Accurate Mass – Time Tag	Monoisotopic Mass (Measured/Calculated)	Δ ppm	Molecular Formula	MS (ESI-Positive)	MS/MS(ESI-Positive) (Varying Collision Energies)	Annotation	References; Further Evidence
BML-LCMS18-20.96-702.1573	702.1573/702.1585	1.67	C36H30O15	[M+H] ⁺ 703.1641 [M+Na] ⁺ 725.1468	10 eV: no fragmentation 50 eV: 541.1118 (100), 421.0543 (48), 393.0621 (10), 271.0578 (4)	Dihydroamentoflavone-hexoside III	
BML-LCMS18-21.15-538.0912	538.0912/538.0900	2.24	C30H18O10	[M+Na] ⁺ 561.0785 [2M+H] ⁺ 1077.1844 [2M+Na] ⁺ 1099.1671	10 eV: no fragmentation 50 eV: 377.0657 (100), 403.0452 (98), 347.0551 (56), 335.0552 (55), 283.0600 (36), 153.0178 (34), 539.0964 (27), 121.0282 (25), 307.0601 (23), 387.0855 (21), 311.0549 (16)	Amentoflavone	Wallace and Markham, 1978; Wang et al., 2015; NMR (Table 2; F10); Authentic Standard
BML-LCMS18-21.41-540.1058	540.1058/540.1057	0.28	C30H20O10	[M+H] ⁺ 541.1130 [M+Na] ⁺ 563.0935	10 eV: no fragmentation 50 eV: 389.1016 (100), 153.0176 (86), 121.0281 (16), 270.0524 (10), 253.0485 (6), 377.0642 (6), 403.0457 (6), 347.0741 (5)	2'' 3''-Dihydro-amentoflavone	Zhang Y.X. et al., 2011
BML-LCMS18-22.03-583.4640	583.4640/583.4640	0.00	C30H18O10	[M+H] ⁺ 539.0979	10 eV: no fragmentation 50 eV: 153.0165 (100), 387.0853 (93), 403.0447 (84), 521.0858 (76), 413.0647 (60), 377.0664 (56), 539.0966 (52), 270.0523 (49), 283.0592 (43), 347.0577 (42), 121.0324 (39), 389.0987 (34), 335.0987 (30), 311.0538 (30)	Robustaflavone	Zhang Y.X. et al., 2011; Wang et al., 2015; NMR (Table 2; F11)
BML-LCMS18-22.59-554.1201	554.1201/554.1213	2.16	C31H22O10	[M+H] ⁺ 555.1270 [M+Na] ⁺ 577.1097 [2M+Na] ⁺ 1133.2517	10 eV: no fragmentation 50 eV: 167.0316 (100), 389.0992 (33), 257.0411 (22), 123.0427 (17), 270.0510 (11)	Dihydro-O-methyl-amentoflavone	Markham, 1984
BML-LCMS18-23.38-552.1052	552.1052/552.1056	0.81	C31H20O10	[M+H] ⁺ 553.1127 [M+Na] ⁺ 575.0941	10 eV: no fragmentation 50 eV: 89.0586 (100), 193.0483 (83), 149.0224 (76), 73.0284 (56), 237.0731 (56), 275.0447 (33), 285.0382 (22), 268.0692 (16), 254.0553 (13), 553.1074 (10), 286.0421 (7), 387.0759 (6)	O-Methyl-amentoflavone (tentative)	Markham, 1984; Wang et al., 2015
BML-LCMS18-23.80-542.1212	542.1212/542.1213	0.18	C30H22O10	[M+H] ⁺ 543.1291 [M+Na] ⁺ 565.1101 [2M+K] ⁺ 1123.1935	10 eV: no fragmentation 50 eV: 153.0180 (100), 271.0592 (75), 391.1183 (26), 147.0427 (21), 297.0376 (10), 166.9972 (10), 179.0342 (10), 423.0729 (5)	Binaringenin	Feuereisen et al., 2017
BML-LCMS18-24.14-540.1060	540.1060/540.1057	0.65	C30H20O10	[M+H] ⁺ 541.1136 [M+Na] ⁺ 563.0941 [2M+Na] ⁺ 1103.1995	10 eV: no fragmentation 50 eV: 153.1777 (100), 389.1015 (96), 257.0438 (53), 270.0514 (33), 285.0385 (18), 421.0533 (17)	Dihydrohinokiflavone (tentative)	
BML-LCMS18-24.50-538.0896	538.0896/538.0900	0.74	C30H18O10	[M+H] ⁺ 539.0979 [M+Na] ⁺ 561.0785 [2M+H] ⁺ 1077.1844	10 eV: no fragmentation 50 eV: 539.0971 (100), 254.0571 (87), 270.0517 (76), 257.0441 (76), 242.0569 (56), 286.0460 (47), 387.0870 (34), 153.0165 (30)	Hinokiflavone	Zhang Y.X. et al., 2011; Wang et al., 2015 NMR (Table 2; F14)

TABLE 2 | Peak annotation based on Spektraris searches with combined HPLC-QTOF-MS and NMR spectroscopy data.

Accurate Mass – Time Tag	Molecular Formula	NMR Solvent	NMR Signals (chemical shift, with integral, signal multiplicity, coupling constant and position in parentheses)	Spektraris Search Score (out of 100)	Annotation	References; Further Evidence
BML-LCMS18-4.66-368.1101	C17H20O9	D ₂ O	7.21 (1H, d, 8 Hz, H-5'), 7.13-7.11 (1H, m, H-4), 6.94 (1H, d, 4 Hz, H-2'), 6.87 (1H, dd 8 and 7 Hz, H-6'), 6.07 (1H, dd, 12 and 4 Hz, H-3), 5.42 (1H, dd, 8 and 4 Hz, H-6), 5.0 (1H, m, H-1') 3.89 (1H, dd, 8 and 4 Hz, H-6), 3.72 (1H, dd, 16 and 4 Hz, H-6), 3.63 (4H, m), 2.64 (2H, m, H-5a,b)	96	3'-Hydroxypsilotin	Balza et al., 1985
BML-LCMS18-5.20-352.1168	C17H20O8	CD ₃ OD	7.4 (1H, d, 4 Hz, H-2), 7.11 (1H, m, H-4), 6.08 (1H, ddd, 16, 8 and 4 Hz, H-3), 5.48 (1H, dd, 12 and 4 Hz, H-6), 3.89 (1H, dd, 12 and 4 Hz, H-6A'), 3.72 (1H, dd, 12 and 4 Hz, H-6B'), 3.5 (4H, m), 2.65 (2H, m, H-5a,b)	90	Psilotin	McInnes et al., 1965
BML-LCMS18-20.73-540.1053	C30H20O10	DMSO-d ₆	7.58 (2H, d, 12 Hz, H-2'' and H-6''), 7.45 (2H, m, H-2' and H-6') 7.04 (1H, d, 8 Hz, H-5'), 6.79 (2H, d, 8 Hz, H-3''' and H-5'''), 6.59 (1H, s, H-3'), 6.33 (1H, s, H-6''), 5.85 (2H, m, H-6 and H-8), 5.47 (1H, d, 4 Hz, H-2), 3.15(1H, m, H-3a), 2.75(1H, m, H-3b)	82	2,3-Dihydro-amentoflavone	Zhang Y.X. et al., 2011
BML-LCMS18-21.15-538.0912	C30H18O10	DMSO-d ₆	8.01 (1H, dd, 8 and 4 Hz, H-6'), 7.94 (1H, d, 4 Hz, H-2'), 7.58 (2H, d, 8 Hz, H-2''' and H-6'''), 7.13 (1H, d, 8 Hz, H-5'), 7.01 (1H, s, H-3), 6.8(1H, s, H-3'), 6.7 (2H, d, 8 Hz, H-3''' and H-5'''), 6.63 (1H, d, 8 Hz, H-8), 6.35 (1H, s, H-6''), 6.21 (1H, d, 4 Hz, H-6)	99	Amentoflavone (3',8''-biapigenin)	Geiger et al., 1993
BML-LCMS18-22.03-583.4640	C30H18O10	DMSO-d ₆	7.9 (3H, m, H-6', H-2''' and H-6'''), 7.2 (1H, d, 4Hz, H-2'), 6.95 (1H, d, 4Hz, H-5'), 6.91 (1H, d, 16 Hz, H-3''' and H-5'''), 6.82(1H, s, H-3). 6.73 (1H, s, H-3'), 6.71 (1H, s, H-8''), 6.43 (1H, d, 4 Hz, H-8), 6.14 (1H, d, 4Hz, H-6)	87	Robustaflavone (3',6''-biapigenin)	Geiger et al., 1993
BML-LCMS18-24.50-538.0896	C30H18O10	DMSO-d ₆	7.9 (3H, m, H-2', H-2''' and H-6'''), 7.18 (2H, d, 4 Hz, H-3' and H-6'), 6.92 (2H, d, 4Hz, H-3''' and H-5''') 6.88 (3H, m, H-3, H-5' and H-3''), 6.75 (1H, s, H-3'), 6.72 (1H, s, H-8''), 6.41 (1H, d, 4 Hz, H-8), 6.13 (1H, d, 4Hz, H-6)	81	Hinokiflavone (4',6''-O-biapigenin)	Geiger et al., 1993

of 40 (arbitrary units); and a step size of 25 μm . Lock mass correction was repeated every 600 s for a duration of 5 s. Other settings: helium gas flow at 90 ml min^{-1} ; trap wave velocity at 311 m s^{-1} ; trap wave height at 4 V; ion mobility wave velocity at 650 m s^{-1} ; ion mobility wave height at 40 V; transfer wave velocity at 191 m s^{-1} ; transfer wave height of 0.1 V; and ion mobility wave delay of 450 μs . The highest signal intensity for the analytes of interest (and thus most desirable signal-to-noise ratio) was achieved in positive polarity for psilotin and 3'-hydroxypsilotin, whereas for amentoflavone negative polarity was preferable. MS/MS experiments were performed by selecting a precursor ion and a collision energy of 30 eV in the transfer cell. MALDI-MS data were processed using the High Definition Imaging software version 1.2 (Waters, Milford, MA, United States) with lock mass correction. Metabolite identification was achieved by comparing accurate mass, MS/MS fragmentation patterns, and ion mobility drift time with those of authentic standards. Signals for isomers of amentoflavone, for example robustaflavone and hinokiflavone, were detectable with authentic standards but not in tissue samples, where their concentrations were too low for MS-based imaging.

RESULTS

Strategy for Multi-Platform Analysis of Metabolites in *P. nudum* Organs

In an attempt to capture chemically diverse metabolites, we used a strategy that accessed five *P. nudum* organs/tissues (below-ground rhizome, above-ground rhizome, and synangia harvested at different developmental stages [green (young), yellow (maturing) and brown (mature)], and generated hydrophilic (methanol/water) and hydrophobic (chloroform/methanol) extracts (Figure 1). These two types of extracts were processed separately for GC-MS analysis (Supplementary Figure S1). Methanol/water extracts were also subjected to HPLC-QTOF-MS analysis in positive ionization mode only (preliminary experiments indicated that chromatographic runs in negative polarity did not add significant spectral information) (Supplementary Figure S1). These data sets were normalized, autoscaled, and then combined for multivariate statistical analyses (Figure 1). Fractions representing selected metabolites of interest were collected from chromatographic separations of extracts and further characterized by $^1\text{H-NMR}$. The different metabolomics platforms (GC-MS and HPLC-QTOF-MS) were chosen because they provide complementary information about different metabolites classes (details presented in the upcoming paragraphs). Cryosections of *P. nudum* above-ground rhizomes were sprayed with a chemical matrix and the cell type-level localization of the most abundant metabolites determined by MALDI-QTOF-MS (Figure 1).

Peak annotation for GC-MS data was achieved based on comparisons of retention times and mass spectral characteristics with those of authentic standards, which led to the high-confidence identification of 83 metabolites in our extracts. MS and MS/MS data from HPLC-QTOF-MS runs (acquired in positive polarity mode) were searched against comprehensive

online databases (MassBank², Metlin³, and National Institute of Standards and Technology⁴). However, these searches were mostly unsuccessful due to a lack of relevant reference spectra in these databases, and we therefore decided to expand the Spektraris online resource⁵ with spectral records acquired as part of this study or extracted from information in the literature. Accurate mass, inferred molecular formula, NMR spectral data, and bibliographic information for 328 metabolites (arylpyrone and biflavonoid aglycones and corresponding glycosides) were integrated into Spektraris-NMR, which now contains spectral records for approximately 21,500 metabolites (status: February, 2019). The combination of accurate mass and retention time data (AMT-tags) acquired by HPLC-QTOF-MS were then searched against Spektraris records (for details of this approach see Cuthbertson et al., 2013), which provided tentative identifications for 27 metabolite peaks (4 annotations with high confidence because of available authentic standards) (Table 1). By also including $^1\text{H-NMR}$ data in Spektraris searches (Fischedick et al., 2015), eight metabolites could be identified with high confidence (Table 2). Relevant structures of arylpyrones and bioflavonoids are shown in Figure 2 and the annotation process for selected peaks is outlined in more detail in the following paragraph.

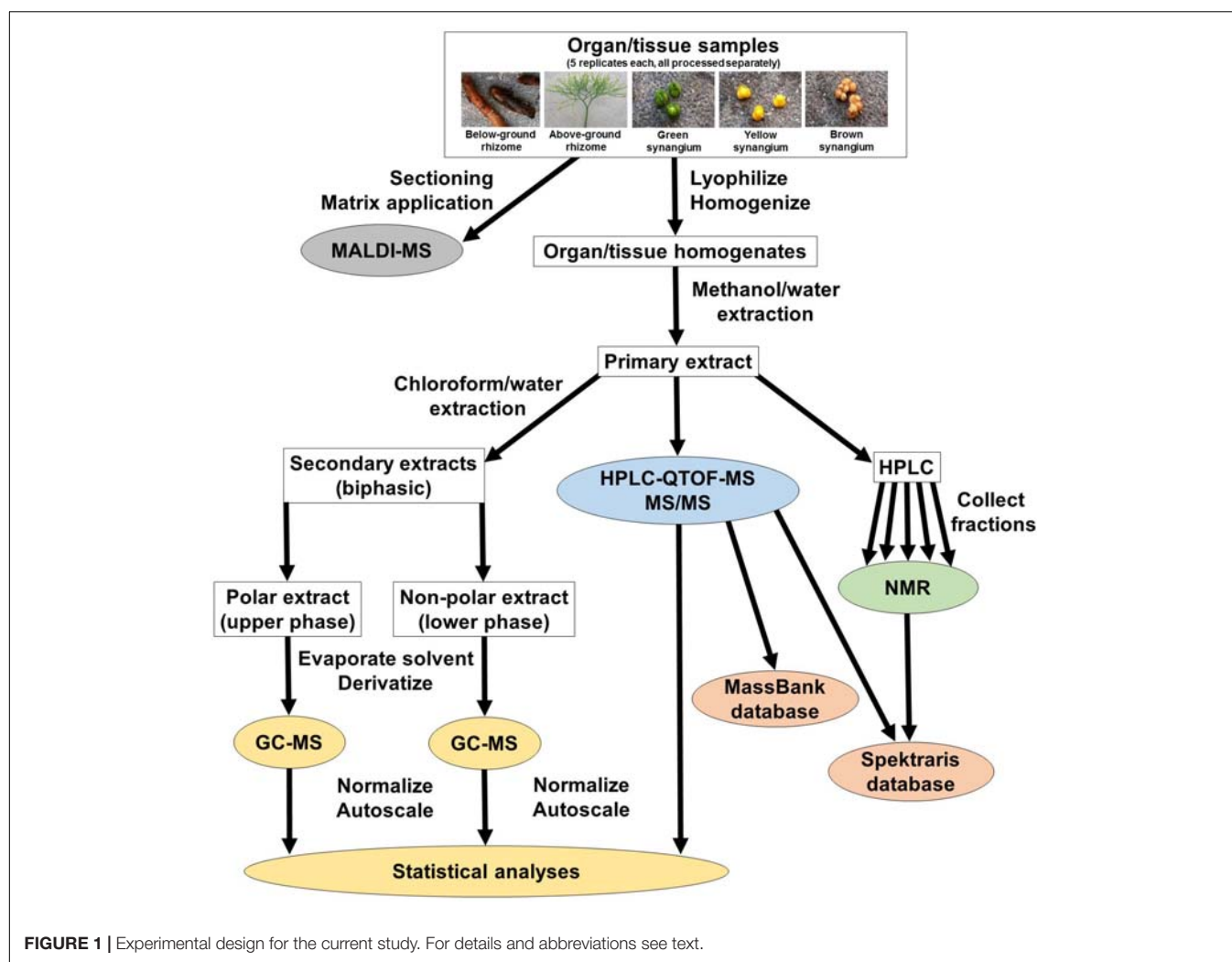
Arylpyrone glycosides with a psilotinin aglycone eluted early (4.6–6.9 min), followed by flavonoid glycosides (11.2–14.6 min), biflavonoid glycosides (14.0–19.3 min), and then biflavonoid aglycones (19.9–24.7 min). In addition to the quasi-molecular ion ($[\text{M}+\text{H}]^+$), adducts ($[\text{M}+\text{Na}]^+$ and $[\text{M}+\text{K}]^+$) and dimers ($[\text{2M}+\text{Na}]^+$ and $[\text{2M}+\text{K}]^+$) were detected consistently for almost all analytes (Table 1). An authentic standard of psilotin (R_t 5.20 min; m/z 353.1250 ($[\text{M}+\text{H}]^+$); $\text{C}_{17}\text{H}_{20}\text{O}_8$) (McInnes et al., 1965) allowed us to investigate the typical fragmentation patterns of this class of arylpyrone glycosides. At low fragmentation energy (10 eV), the loss of glucose generated a fragment representing the psilotinin aglycone (m/z 191.0699; $[\text{M}-\text{Glc}]^+$; $\text{C}_{11}\text{H}_{10}\text{O}_3$), a second fragment with additional loss of water (m/z 173.0598; $[\text{M}-\text{Glc}-\text{H}_2\text{O}]^+$; $\text{C}_{11}\text{H}_{10}\text{O}_3$), a third fragment at m/z 123.0438 ($\text{C}_7\text{H}_7\text{O}_2$), and a fourth fragment with m/z 107.0487 ($\text{C}_7\text{H}_7\text{O}$). The absorption spectrum of the peak in chromatograms and the authentic psilotin standard was essentially identical (Supplementary Figure S2). Furthermore, following purification of the peak by HPLC, the NMR spectrum of the isolated metabolite matched literature reports (McInnes et al., 1965) (Table 2). The signature fragments generated from the psilotin HPLC-QTOF-MS peak (corresponding to $\text{C}_{11}\text{H}_{10}\text{O}_3$ and $\text{C}_7\text{H}_7\text{O}$) were also detected in MS/MS spectra of four additional peaks (plus further common fragments at 50 eV). The molecular ions of two of these peaks indicated the potential presence of two hexose moieties (R_t 5.12 and 6.83 min; m/z 515.1751 ($[\text{M}+\text{H}]^+$); $\text{C}_{23}\text{H}_{30}\text{O}_{13}$), which was corroborated by fragmentation patterns ($[\text{M}-2 \times \text{Glc}]^+$ and $[\text{M}-2 \times \text{Glc}-\text{H}_2\text{O}]^+$). The third peak of this series had a molecular ion consistent with 3'-hydroxypsilotin (R_t 4.66 min; m/z 369.1198

²<https://massbank.eu/MassBank/>

³<https://metlin.scripps.edu/>

⁴<https://chemdata.nist.gov/>

⁵<http://langelabtools.wsu.edu/spektraris/>



($[M+H]^+$; $C_{17}H_{20}O_9$) (Balza et al., 1985), an annotation that was confirmed based on the NMR spectrum of the isolated metabolite (Tables 1, 2). The fourth peak appeared to contain the same aglycone but with two attached hexose moieties and was therefore tentatively annotated as 3'-hydroxypsilotinin-di-O-hexoside (R_t 4.76 min; m/z 553.1530 ($[M+Na]^+$; $C_{23}H_{30}O_{14}$).

Based on the characteristics of the peak corresponding to apigenin 7-O-glucoside, for which an authentic standard was available (R_t 14.26 min; m/z 433.1133 ($[M+H]^+$; $C_{21}H_{20}O_{10}$) (Wallace and Markham, 1978), the fragment indicative of an apigenin aglycone was m/z 271.0597 ($[M-Glc]^+$; $C_{15}H_{10}O_5$), with m/z 153.0175 ($C_7H_5O_4$) representing a second prominent fragment obtained from this flavone aglycone (Kachlicki et al., 2016). Two additional peaks had comparable fragmentation patterns, one of which showed a quasi-molecular ion corresponding to apigenin-7-O-rhamnoglucoside (R_t 14.50 min; m/z 579.1712 ($[M+H]^+$; $C_{27}H_{30}O_{14}$), a metabolite that had previously been reported to occur in *P. nudum* (Wallace and Markham, 1978). The mass spectrum of the second of these peaks was indicative of a metabolite with two hexose moieties (R_t 11.20 min; m/z 595.1654 ($[M+H]^+$; $C_{27}H_{30}O_{15}$) and therefore

likely corresponds to an apigenin di-hexoside. Interestingly, apigenin-6,8-di-C-glucoside (vicenin-2) was described before as an abundant constituent of *P. nudum* extracts (Markham, 1984), which was used as a tentative annotation for the peak of interest (Table 1).

All biflavonoid glycosides thus far characterized from *P. nudum* extracts contained an amentoflavone (3',8''-biapigenin) aglycone with likely O-linked hexose moieties (Wallace and Markham, 1978; Markham, 1984). The amentoflavone authentic standard ($C_{30}H_{18}O_{10}$) eluted at 21.15 min (Table 1). Six additional peaks with the characteristic m/z 539.0963 fragment (corresponding to this aglycone) plus common MS/MS fragments of the aglycone (Zhang Y.X. et al., 2011; Feureisen et al., 2017) were detected in our extracts. Based on their quasi-molecular ion, these peaks were tentatively annotated as amentoflavone-hexosides (R_t 18.12 and 19.16 min; m/z 701.1490 ($[M+H]^+$; $C_{36}H_{28}O_{15}$), amentoflavone-di-hexosides (R_t 15.78 and 16.60 min; m/z 863.2021 ($[M+H]^+$; $C_{42}H_{38}O_{20}$) or amentoflavone-tri-hexosides (R_t 13.99 and 14.82 min; m/z 1025.2551 ($[M+H]^+$; $C_{48}H_{48}O_{25}$). Three peaks in the same retention

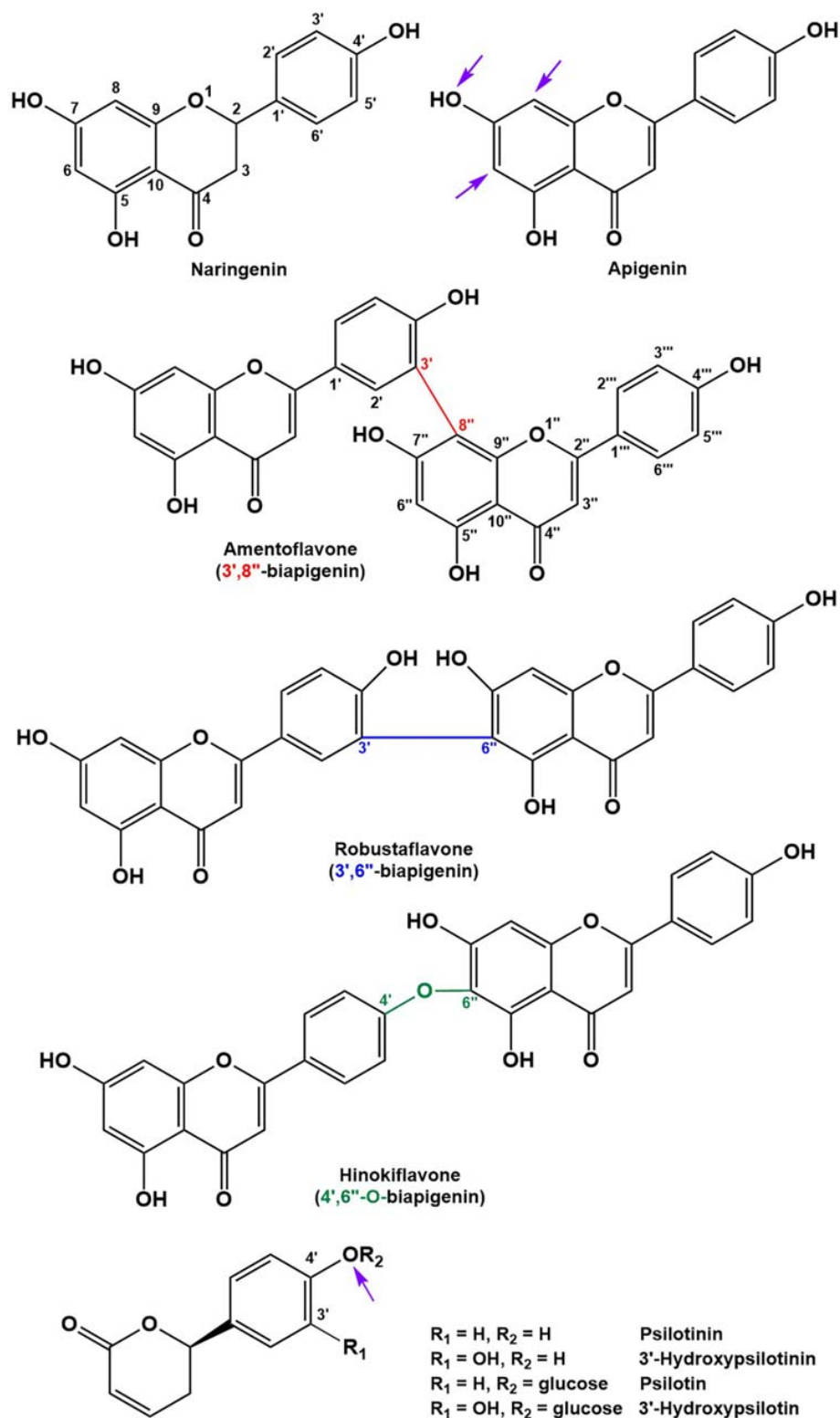


FIGURE 2 | Structures of the main flavonoid, biflavonoid and arylpyrone aglycones of *P. nudum* (positions for common glycosylation sites are indicated with purple arrows).

time region (R_t 17.56, 18.28, and 20.96 min) had a quasi-molecular ion (m/z 703.1642 ($[M+H]^+$); $C_{36}H_{30}O_{15}$) and characteristic aglycone fragment (m/z 541.0545; $C_{30}H_{20}O_{10}$) indicative of two additional mass units compared to amentoflavone-hexosides. The aglycone in these cases is dihydroamentoflavone (C–C-linked dimer of apigenin and naringenin), and the peaks were therefore annotated as dihydroamentoflavone-hexosides (Table 1). The absorption spectrum of the amentoflavone standard in the ultraviolet and visible range was identical to that of the corresponding peak in our HPLC runs (Supplementary Figure S2). Two arylpyrone aglycones, 3'-hydroxypsilotinin (R_t 5.52; m/z 207.0652 ($[M+H]^+$); $C_{11}H_{10}O_4$) and psilotinin (R_t 8.31; m/z 191.0703 ($[M+H]^+$); $C_{11}H_{10}O_3$), were tentatively identified based on chromatographic properties and similarity of MS/MS fragmentation patterns to those of their glycosides (Supplementary Figure S3).

To enable the differentiation of biflavonoid aglycones, four fractions collected by HPLC were subjected to 1H -NMR spectroscopy. Taking into account the previously published elution order of biflavonoids on reversed-phase HPLC materials (Zhang Y.X. et al., 2011), MS and MS/MS data, and by combining this information with MS and NMR data searches against the Spektraris database, four peaks could be identified with high confidence (20.73 min, 2,3-dihydroamentoflavone; 21.15 min, amentoflavone; 22.03 min, robustaflavone; and 24.50 min, hinokiflavone (Figure 2 and Tables 1, 2). The peak at 19.96 min (m/z 555.0931 ($[M+H]^+$); $C_{30}H_{18}O_{11}$) was tentatively identified as hydroxyamentoflavone based on its earlier elution (compared to amentoflavone) and MS/MS fragmentation patterns (Zhang Y.X. et al., 2011). Analogous comparisons allowed the tentative identification of dihydro-*O*-methyl-amentoflavone (R_t 22.59 min; m/z 555.1270 ($[M+H]^+$); $C_{31}H_{22}O_{10}$), *O*-methyl-amentoflavone (R_t 23.38 min; m/z 553.1127 ($[M+H]^+$); $C_{31}H_{20}O_{10}$), binaringenin (R_t 23.80 min; m/z 543.1291 ($[M+H]^+$); $C_{30}H_{22}O_{10}$), and dihydrohinokiflavone (R_t 24.14 min; m/z 541.1136 ($[M+H]^+$); $C_{30}H_{20}O_{10}$) (Markham, 1984; Zhang Y.X. et al., 2011; Wang et al., 2015; Feuereisen et al., 2017) (Table 1).

A third class of metabolites with high abundance in HPLC–QTOF–MS runs had the chromatographic and mass spectral properties of highly functionalized triterpenoids (steroids) (Supplementary Table S2). While the typical membrane sterols of *P. nudum* have been reported before (Akihisa et al., 1992), the more functionalized steroids detected here have not been mentioned in previous studies. A more detailed characterization of these underexplored specialized metabolites will be the subject of future endeavors to further evaluate chemicals diversity in the fern lineage.

Principal Component Analysis Differentiates Metabolomics Data Sets From Different *P. nudum* Organs

Multivariate statistical analyses, such as PCA, aid with reducing the complexity of extensive data sets into a smaller number of Principal Components (PCs). When this approach was

brought to bear on our combined GC–MS and HPLC–QTOF–MS data sets, the first three PCs accounted for roughly 71% of the varied influences of the original characteristics (metabolite patterns across all sample types), and indicated a clear separation of the five sample types (*P. nudum* below-ground rhizome, above-ground rhizome, and synangia harvested at three different developmental stages), with a tight clustering of biological replicates (Figures 3A,B). Below-ground rhizome samples were characterized by positive scores in PC1 and PC3, with neutral values in PC2. Above-ground rhizome samples also had positive scores in PC1 but negative scores in PC2 and PC3 (Figures 3A,B). All samples from synangia had negative scores in PC1, but were differentiated by a combination of negative PC2/positive PC3 scores (green synangia), negative PC2/PC3 scores (yellow synangia) or positive PC2/neutral PC3 scores (brown synangia) (Figures 3A,B).

Component loadings were then evaluated for characteristics that contributed to the differences among sample clusters in PCA and visualized in a biplot (Figure 3C and Supplementary Table S3). The scores for below-ground rhizome samples (positive PC1 and PC2) corresponded to the vectors for carbohydrate monomers and dimers (e.g., glucose, fructose, galactose, and raffinose) and small organic acids (malic acid, citric acid, and succinic acid). Above-ground rhizome samples occupied a biplot position (positive PC1, negative PC2) closer to the direction of vectors for some arylpyrone glycosides (psilotin and psilotinin-di-O-hexoside II) and sucrose (which had high concentrations in both above- and below-ground rhizomes) (Figure 3C). The unique position of brown synangia (negative PC1/positive PC2) correlated with the vector for biflavonoid glycosides (e.g., dihydroamentoflavone hexoside I and amentoflavone-tri-O-hexoside II). Scores for green and yellow synangia were similar (negative scores in both PC1 and PC2) and correlated with the direction of vectors for arylpyrone glycosides (e.g., psilotinin-di-O-hexoside I) and biflavonoid aglycones (e.g., amentoflavone and hinokiflavone) (Figure 3C).

Organ-Specific Accumulation of Metabolites

The PCA component loadings indicated that specific metabolite classes might explain the separation of sample types. We therefore generated a heatmap of metabolite accumulation patterns across *P. nudum* organs (Figure 4). The relative quantities of five biflavonoid glycosides, based on normalized peak areas, were quite high in brown synangia, followed by yellow and green synangia. These metabolites were also of fairly high abundance in samples of above-ground rhizomes, but extremely low in below-ground rhizomes (Figure 4). The quantities of six additional biflavonoid glycosides were considerably lower in all samples. Amentoflavone was the by far most abundant biflavonoid aglycone, with very high amounts present in yellow and brown synangia, relatively high quantities in above-ground rhizomes and green synangia, and fairly low levels in below-ground rhizomes (Figure 4). Similar patterns were observed for five additional biflavonoid

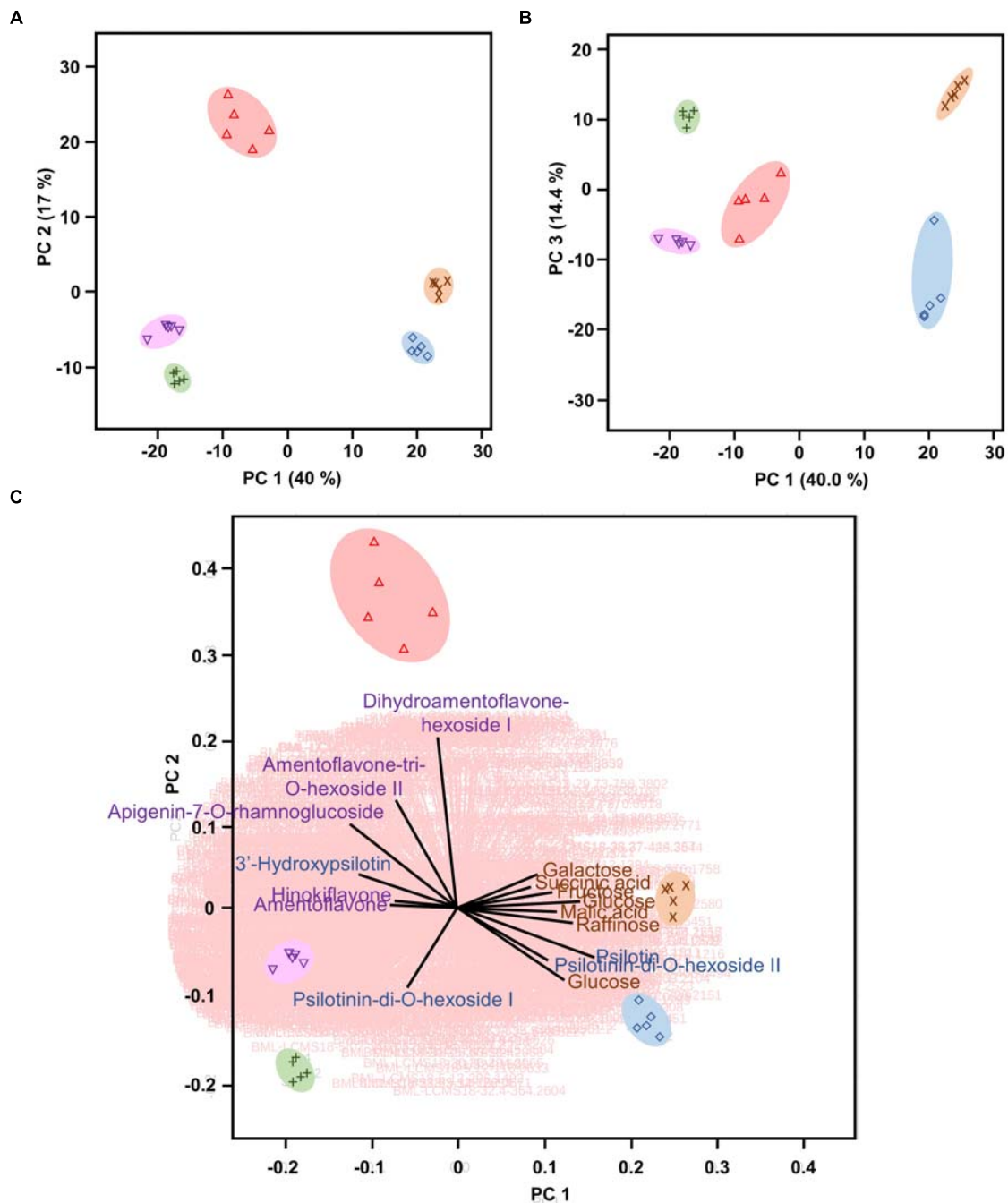


FIGURE 3 | *Psilotum nudum* organs have a unique metabolic fingerprint, based on PCA of combined GC-MS and HPLC-QTOF-MS data. **(A)** PCA plot for PC1 and PC2. **(B)** PCA plot for PC1 and PC3. **(C)** Bi-plot with selected component loadings (metabolites) highlighted (the vectors of the remaining AMT tags are shown with red lines in the background). The full component loadings are provided in **Supplementary Table S3**. Symbols: rhizome (below-ground), purple cross; rhizome (above-ground), blue diamond; green syngonium, green plus sign; yellow syngonium, pink inverted triangle; brown syngonium, red triangle.

aglycones (three dihydroapiogenin isomers, binaringenin and hinokiflavone), albeit at much lower abundance compared to amentoflavone. Among arylpyrone glycosides, psilotin was most abundant in rhizomes and above-ground rhizomes, but was also accumulated to appreciable amounts in syngonia (Figure 4). 3'-Hydroxypsilotin was primarily found in syngonia, with an

abundance comparable to that of psilotin. Three other arylpyrone glycosides were of relatively low abundance in all samples.

Sucrose was equally abundant in rhizomes, above-ground rhizomes and green syngonia (Figure 4). Glucose, fructose and other small molecule carbohydrates were most abundant in rhizomes, with significantly lower amounts being present in all

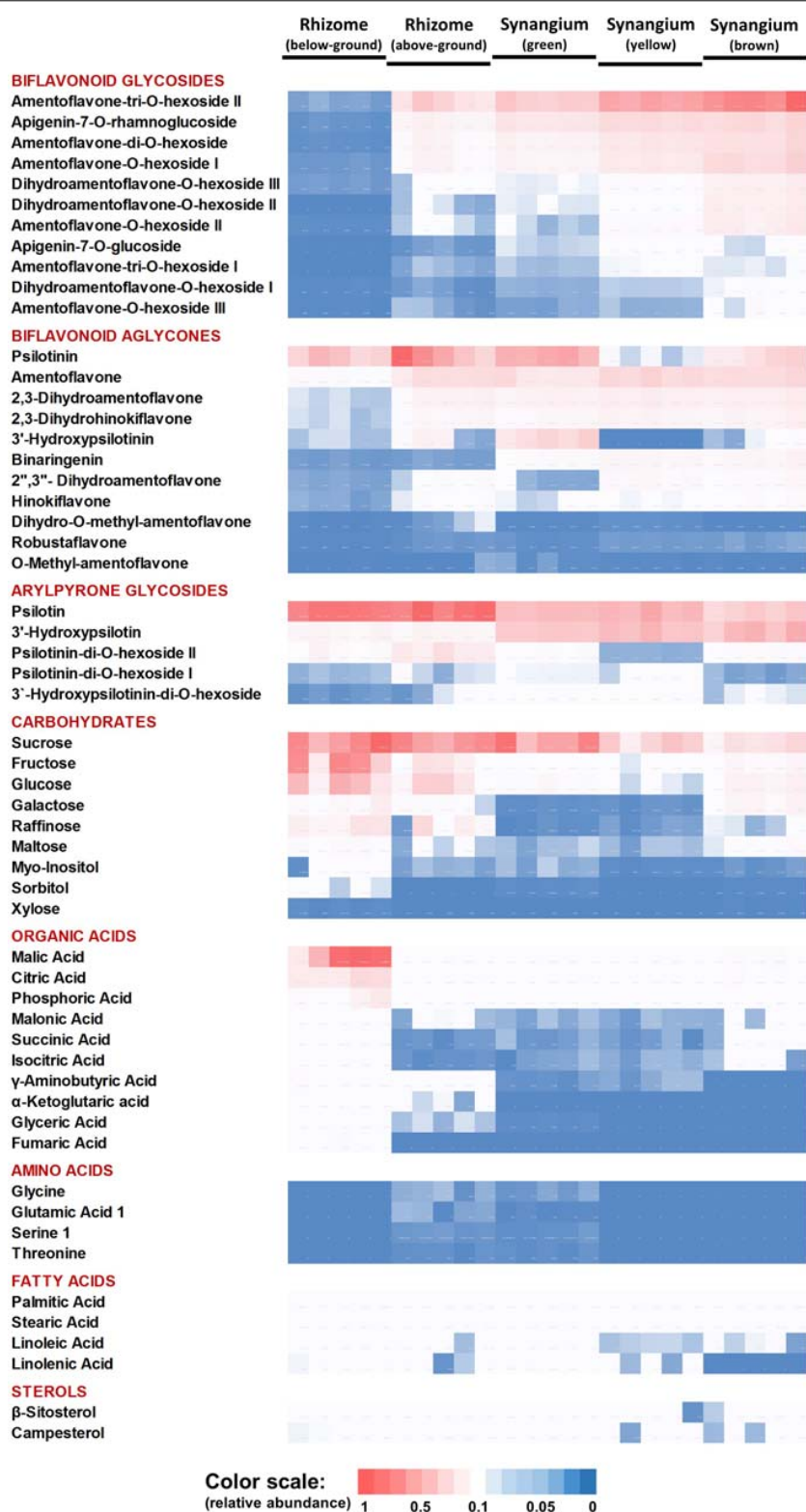


FIGURE 4 | Heat map visualizing the organ-specificity of metabolite accumulation in *P. nudum*. A color code (red-white-blue) is used to indicate the relative abundance of metabolites based on normalized mass spectral intensities.

other samples. The highest levels of small organic acids were also found in rhizomes. While malic and citric acid were fairly abundant in all samples, other organic acids (e.g., α -ketoglutaric acid, glyceric acid and fumaric acid) were detected at considerably lower levels in rhizomes and yet lower levels in all other samples (Figure 4).

MALDI-MS Imaging Indicates Preferential Accumulation of Amentoflavone and Arylpyrone Glycosides in Stem Epidermis and Outer Cortex

Building on recent successes with MS-based imaging of sesquiterpene alkaloids and triterpenoids (Lange et al., 2017), MALDI-MS was employed for localizing metabolites of interest in the current study. Two arylpyrone glycosides, psilotin and 3'-hydroxypsilotin, and a biflavonoid aglycone, amentoflavone, were selected for because they were highly abundant in tissue samples (MS-based imaging is much less sensitive compared to tissue extraction followed by HPLC-QTOF-MS) and were available as authentic standards in sufficient quantities for methods development. Based on the results of preliminary experiments, 30 μ m cryosections of above-ground rhizomes served as biological material, 2,5-dihydroxybenzoic acid was chosen as matrix substance to aid with ionization of metabolites desorbed from tissue sections, and leucine-enkephalin was selected to provide an external lock mass. The ionization of psilotin and 3'-hydroxypsilotin was most effective in positive ionization mode, where potassium adducts (m/z 391.0797 and 407.0750, respectively) were readily detectable with unique drift times in the ion mobility cell. Mass spectrometric signals for psilotin, 3'-hydroxypsilotin and amentoflavone were highest in the epidermal and outer cortex layers, which collectively form the chlorenchyma (Figures 5A–D). The 3'-hydroxypsilotin signal was also apparent, albeit at significantly lower abundance, in the protostele. Amentoflavone ionized particularly well in negative mode, with the quasi-molecular ion being more abundant than adducts (m/z 537.0827) and traveling through the ion mobility cell with a unique drift time (Figures 5B,C,E). Based on MALDI-MS experiments performed with above-ground rhizome extracts, the normalized peak area for amentoflavone was 5-fold higher than that of hinokiflavone and 47-fold higher than that of robustaflavone, and the abundance of the latter two metabolites was too low for localization studies.

DISCUSSION

Expanding the Coverage of Spectral Databases to Incorporate Information on Chemical Diversity in the Fern Lineage

Biflavonoids have long been known to accumulate prominently across the bryophytes, pteridophytes and gymnosperms, with only sporadic occurrence in the angiosperms (Geiger and Quinn, 1988; Iwashina, 2000). When we began processing the data presented as part of the current study with *P. nudum*, we noticed

a surprising paucity of spectral data relating to biflavonoids in publicly available MS and NMR databases. We therefore embarked on a literature search to gather phytochemical and spectral data for this important class of metabolites, which was then used to generate 328 new spectral records for the Spektraris online resource (Cuthbertson et al., 2013; Fishedick et al., 2015). Additionally, electronic files representing the MS/MS data acquired with biflavonoids were submitted to MassBank, a widely used online mass spectral repository (Horai et al., 2010).

The orthogonal data sets acquired in this study (retention time on GC or HPLC, quasi-molecular ion (and inferred molecular formula), MS/MS data, and NMR spectra), combined with the use of authentic standards, aided substantially in peak annotation. The inclusion of NMR data was particularly impactful for the annotation of peaks for the biflavonoids (amentoflavone, robustaflavone, and hinokiflavone) that consist of two fused apigenin molecules (differing only in the coupling position). Using our integrative approach, a total of 83 GC-MS and 8 HPLC-QTOF-MS peaks were identified with very high confidence. An additional 23 HPLC-QTOF-MS peaks were tentatively identified (for example, amentoflavone-tri-O-hexoside I, where uncertainty pertains only to the position and exact nature of the hexose moiety) (Table 1). While we were able to determine the structures of some of the more abundant aglycones, the identification of biflavonoid glycosides, which occur as larger families of closely related structures, has proven much more difficult. Our data sets also contained a very large number of peaks that could not be identified. Some of these, based on peak area counts, appeared to be fairly abundant. These results indicate that significant efforts will be needed to generate a more comprehensive account of chemical diversity in *P. nudum* and, more broadly, in the fern lineage.

Below-Ground Rhizome of *P. nudum* Contains High Levels of Soluble Sugars and Organic Acids, Possibly Indicating Differential Nutrient Allocation

The above- and below-ground portions of the *P. nudum* rhizome are part of the same organ and it is thus notable that, in our study, significantly higher amounts of soluble sugars (in order of abundance: fructose, glucose, raffinose and galactose) and organic acids (in order of abundance: malic acid, citric acid and phosphoric acid) were present in the below-ground part of the rhizome. The abundance of soluble sugars might be interpreted as evidence for a storage function for *P. nudum* rhizomes but, to the best of our knowledge, the corresponding storage sugar polymers have not been analyzed in this species. The chemical properties of rhizome starches have been reported for other ferns (Zhang S. et al., 2011; Yu et al., 2015) and this work indicates indirectly (based on the high abundance of sugar precursors) that storage function is a possibility. It is also conceivable that relatively high levels of soluble sugars and organic acids are a reflection of active metabolism to support horizontal rhizome growth in *P. nudum*. However, while information is available regarding the correlation of fern development and some classes of metabolites (White and Turner, 1995; Abul et al., 2010), we were not able to find literature

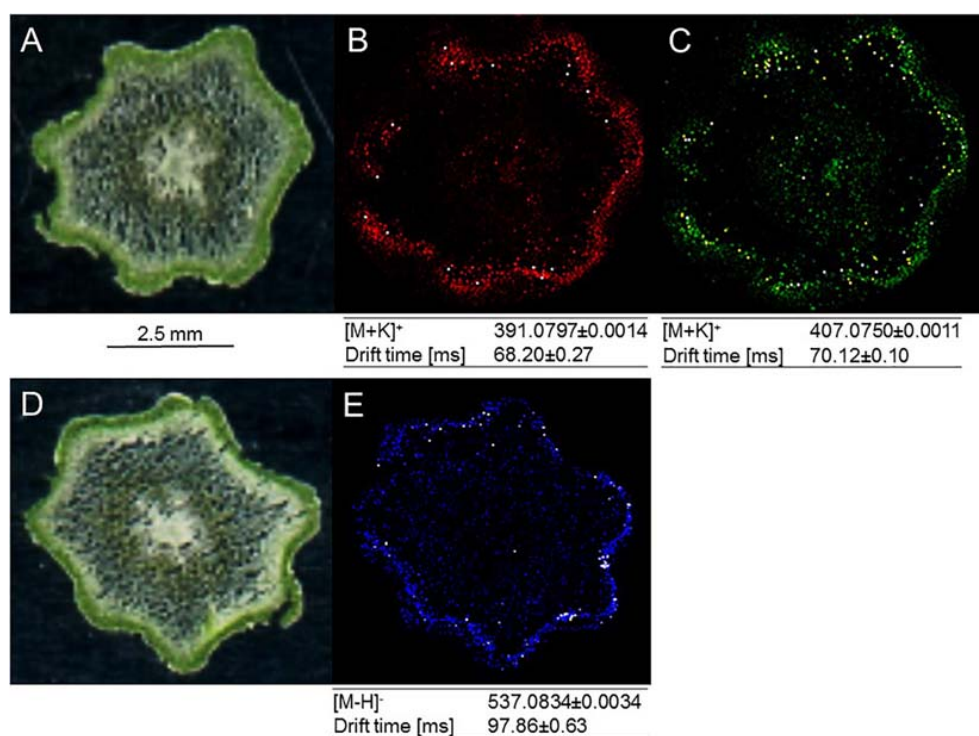


FIGURE 5 | MALDI-MS imaging of specialized metabolites in cross sections of *P. nudum* above-ground rhizomes (stems). **(A)** Image of specimen for arylpyrone glycoside analysis; **(B)** [M+K]⁺ signal for psilotin (colored in red); **(C)** [M+K]⁺ signal for 3'-hydroxypsilotin (colored in green); **(D)** image of specimen for biflavonoid analysis; **(E)** [M-H]⁻ signal for amentoflavone (colored in blue). Note the accumulation of metabolites of interest in the chlorenchyma (dark green tissue in specimen images).

on soluble sugar quantities in fern rhizomes. Further research is clearly necessary to begin to appreciate the tissue specialization within fern rhizomes.

Rhizomes Accumulate Particularly High Amounts of Psilotin, an Arylpyrone Glycoside With Demonstrated Biological Activities

Our data indicated that psilotin and psilotinin, both arylpyrones unique to the Psilotaceae, were most abundant in the rhizome (both below- and above-ground), while being only half or one-third as abundant in samples from syngangia. The biflavonoid amentoflavone was also highly abundant in the above-ground part of the rhizome but occurred at fairly low quantities in the below-ground parts (**Figure 4**). This begs the question if psilotin and its aglycone psilotinin might play a particular role in the below-ground rhizome, where arylpyrones are major constituents. Interestingly, it was demonstrated more than 40 years ago that psilotin acts as a germination inhibitor for turnip, onion and lettuce seeds (Siegel, 1976). It is, therefore, conceivable that psilotin (and possibly its aglycone as well) plays a defensive or allelochemical role in and around the below-ground rhizome. Psilotin was also shown to have antifeedant activities against the European corn borer (*Ostrinia nubilalis*) at concentrations below those present in *P. nudum*

(Arnason et al., 1986). However, in the absence of more complete data on the bioactivities of arylpyrones, this interpretation is highly speculative. It is also unknown how psilotin might be secreted into the rhizosphere to exert allelochemical activities. The fact that the inhibitory effects of psilotin on germination can be reversed by the addition of GA₃ (Siegel, 1976), a gibberellin hormone, can be interpreted as evidence for a possible role of this arylpyrone in growth regulation, but the mechanism and target(s) of such an activity have not yet been explored. In the above-ground rhizome, psilotin and amentoflavone (the latter also exerting high bioactivity; Yu et al., 2017) may act collectively as defense metabolites. Currently, information about such activities has been inferred from *in vitro* assays only and it would thus be informative to also assess potential defensive functions of arylpyrones and bioflavonoids in *in vivo* investigations.

Occurrence of Biflavonoids and Arylpyrones in Chlorenchyma Is Consistent With Function as Sunscreen Pigments

Based on our MALDI-MS imaging data, psilotin and amentoflavone are accumulated preferentially in the photosynthetically active tissues of above-ground rhizomes (above-ground rhizomes) (**Figure 5**). Considering the absorption

characteristics of these metabolites (Supplementary Figure S2), a protective function against excess photosynthetically active radiation and certain wavelengths (e.g., high energy ultraviolet-B) radiation would be a reasonable hypothesis for their tissue-level localization (Yamaguchi et al., 2009; Waterman et al., 2017). Our localization data sets for amentoflavone (chlorenchyma) are also consistent with the literature for other plants. For example, amentoflavone was accumulated preferentially in the leaf epidermis in *Agathis robusta* (Gadek et al., 1984) and *Ginkgo biloba* (Beck and Stengel, 2016). An interesting, as yet unanswered, question pertains to the functional role of the differential subcellular localization one would predict for the metabolites of interest. Psilotin is likely stored in the vacuole, in analogy to other (polar) phenolic glycosides (Wink, 1993), while amentoflavone is an apolar biflavonoid aglycone that was previously found to be associated with cell walls (Gadek et al., 1984). Both locations allow for the sequestration of these bioactive metabolites, thereby protecting cellular metabolism in different subcellular locations (Agapakis et al., 2012). Another advantage of the differential localization of psilotin and amentoflavone could be that greater quantities of these pigments can be accumulated, but this hypothesis remains to be tested.

DATA AVAILABILITY

All datasets generated for this study are included in the manuscript and/or the **Supplementary Files**.

AUTHOR CONTRIBUTIONS

BL conceived the work and wrote the manuscript, with input from all authors. DŠ, VP, NS, and BL designed the experiments and analyzed the data. MW served as mentor for VP. VP generated GC-MS data. DŠ obtained HPLC-QTOF-MS and MALDI-MS data. NS produced and interpreted NMR data and contributed to the generation of new spectral records for the Spektraris online resource.

REFERENCES

- Abul, Y., Menéndez, V., Gómez-Campo, C., Revilla, M. A., Lafont, F., and Fernández, H. (2010). Occurrence of plant growth regulators in *Psilotum nudum*. *J. Plant Physiol.* 167, 1211–1213. doi: 10.1016/j.jplph.2010.03.015
- Agapakis, C. M., Boyle, P. M., and Silver, P. A. (2012). Natural strategies for the spatial optimization of metabolism in synthetic biology. *Nat. Chem. Biol.* 8, 527–535. doi: 10.1038/nchembio.975
- Akihisa, T., Kawashima, T., Takahashi, S., Sahashi, N., Okamoto, T., Niiya, I., et al. (1992). Sterols and fatty acids of a whisk fern *Psilotum nudum*. *J. Am. Oil Chem. Soc.* 69, 1232–1235. doi: 10.1007/BF02637687
- Arnason, J. T., Philogène, J. R., Donskov, N., Muir, A., and Towers, G. H. N. (1986). Psilotin, an insect feeding deterrent and growth reducer from *Psilotum nudum*. *Biochem. Syst. Ecol.* 14, 287–289. doi: 10.1016/0305-1978(86)90103-1
- Balza, F., Muir, A. D., and Towers, G. H. N. (1985). 3'-Hydroxypsilotin, a minor arylpyrone glycoside from *Psilotum nudum*. *Phytochemistry* 24, 529–531. doi: 10.1016/S0031-9422(00)80761-X
- Banks, H. P. (1975). Reclassification of the psilophyta. *Taxon* 24, 401–413.

FUNDING

This work was supported in part by seed funds from the USDA National Institute of Food and Agriculture, Hatch Umbrella project #1015621. DŠ acknowledges financial support by the European Union's Seventh Framework Programme under grant agreement # 291823, Marie Curie FP7-PEOPLE-2011-COFUND NEWFELPRO, project 64.

ACKNOWLEDGMENTS

We would like to thank the Institute of Biological Chemistry's greenhouse staff, Ms. Julie Thayer and Mr. Devon Thrasher, for plant maintenance. We are also indebted to Dr. Jordan Zager for establishing a pipeline for PCA. Ms. Ingrid Wokey and Mr. Nicholas Elms are acknowledged for their work on generating Spektraris records.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.00868/full#supplementary-material>

FIGURE S1 | Representative GC-MS and HPLC-QTOF-MS chromatograms (total ion current) of *P. nudum* extracts.

FIGURE S2 | Absorbance spectra of peaks detected in HPLC-QTOF-MS runs and comparison with those of authentic standards.

FIGURE S3 | MS/MS spectra (acquired at 30 eV collision energy) of psilotin and psilotinin.

TABLE S1 | Peak identification in GC-MS runs.

TABLE S2 | Normalized peak areas for GC-MS and HPLC-QTOF-MS data.

TABLE S3 | Component loadings of a PCA analysis with combined GC-MS and HPLC-QTOF-MS data.

TABLE S4 | NMR acquisition parameters.

TABLE S5 | Bioflavonoids and related metabolites for which new Spektraris spectral records were generated.

- Beck, S., and Stengel, J. (2016). Mass spectrometric imaging of flavonoid glycosides and biflavonoids in *Ginkgo biloba* L. *Phytochemistry* 130, 201–206. doi: 10.1016/j.phytochem.2016.05.005
- Bino, R. J., Hall, R. D., Fiehn, O., Kopka, J., Saito, K., Draper, J., et al. (2004). Potential of metabolomics as a functional genomics tool. *Trends Plant Sci.* 9, 418–425. doi: 10.1016/j.tplants.2004.07.004
- Chiu, P. L., Patterson, G. W., and Salt, T. A. (1988). Sterol composition of pteridophytes. *Phytochemistry* 27, 819–822. doi: 10.1016/0031-9422(88)84099-8
- Cooper-Driver, G. (1977). Chemical evidence for separating the Psilotaceae from the Filicales. *Science* 198, 1260–1262. doi: 10.1126/science.198.4323.1260
- Cuthbertson, D., Andrews, P. K., Reganold, J. P., Davies, N. M., and Lange, B. M. (2013). Utility of metabolomics toward assessing the metabolic basis of quality traits in apple fruit with an emphasis on antioxidants. *J. Agric. Food Chem.* 60, 8552–8560. doi: 10.1021/jf3031088
- Doyle, J. A. (2018). Phylogenetic analyses and morphological innovations in land plants. *Annu. Plant Rev.* 45, 1–50. doi: 10.1002/9781119312994.apr0486

- Ducket, J. G., and Ligrone, R. (2005). A comparative cytological analysis of fungal endophytes in the sporophyte rhizomes and vascularized gametophytes of *Tmesipteris* and *Psilotum*. *Can. J. Bot.* 83, 1443–1456. doi: 10.1139/b05-102
- Feuereisen, M. M., Gamero Barraza, M., Zimmermann, B. F., Schieber, A., and Schulze-Kaysers, N. (2017). Pressurized liquid extraction of anthocyanins and biflavonoids from *Schinus terebinthifolius* Raddi: a multivariate optimization. *Food Chem.* 214, 564–571. doi: 10.1016/j.foodchem.2016.07.002
- Fiehn, O., Sumner, L. W., Rhee, S. Y., Ward, J., Dickerson, J., Lange, B. M., et al. (2007). Minimum reporting standards for plant biology context information in metabolomic studies. *Metabolomics* 3, 195–201. doi: 10.1007/s11306-007-0068-0
- Fischedick, J. T., Johnson, S. R., Ketchum, R. E., Croteau, R. B., and Lange, B. M. (2015). NMR spectroscopic search module for Spektraris, an online resource for plant natural product identification taxane diterpenoids from *Taxus* × media cell suspension cultures as a case study. *Phytochemistry* 113, 87–95. doi: 10.1016/j.phytochem.2014.11.020
- Gadek, P. A., Quinn, C. J., and Ashford, A. E. (1984). Localization of the biflavonoid fraction in plant leaves, with special reference to *Agathis robusta* (C. Moore ex F. Muell.) *F.m. Bail. Austr. J. Bot.* 32, 15–31. doi: 10.1071/BT9840015
- Geiger, H., and Quinn, C. (1988). “Biflavonoids,” in *The Flavonoids. Advances in Research Since 1980*, ed. J. B. Harborne (London: Chapman and Hall), 99–124. doi: 10.1007/978-1-4899-2913-6_4
- Geiger, H., Seeger, T., Hahn, H., Zinsmeister, D., Markham, K., and Wong, H. (1993). 1H NMR assignments in biflavonoid spectra by proton-detected C-H correlation. *Z. Naturforsch.* 48c, 821–826. doi: 10.1515/znc-1993-11-1201
- Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., et al. (2010). MassBank: a public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.* 45, 703–714. doi: 10.1002/jms.1777
- Iwashina, T. (2000). The structure and distribution of the flavonoids in plants. *J. Plant Sci.* 113, 287–299. doi: 10.1007/PL00013940
- Kachlicki, P., Piasecka, A., Stobiecki, M., and Marczak, Ł (2016). Structural characterization of flavonoid glycoconjugates and their derivatives with mass spectrometric techniques. *Molecules* 21:1494. doi: 10.3390/molecules21111494
- Kenrick, P., and Crane, P. R. (1997). The origin and early evolution of plants on land. *Nature* 389, 33–39. doi: 10.1038/37918
- Lange, B. M., Fischedick, J. T., Lange, M. F., Srividya, N., Šamec, D., and Poirier, B. C. (2017). Integrative approaches for the identification and localization of specialized metabolites in *Tripterium* roots. *Plant Physiol.* 173, 456–469. doi: 10.1104/pp.15.01593
- Markham, K. R. (1984). The structures of amentoflavone glycosides isolated from *Psilotum nudum*. *Phytochemistry* 23, 2053–2056. doi: 10.1016/S0031-9422(00)84969-9
- McInnes, A. G., Yoshida, S., and Towers, G. H. N. (1965). A arylpyrone glycoside from *Psilotum nudum* (L.) Griseb. *Tetrahedron* 21, 2939–2946. doi: 10.1016/S0040-4020(01)98380-2
- Nilsen, E. T. (1995). “Stem photosynthesis: extent, patterns, and role in plant carbon economy,” in *Plant Above-Ground Rhizomes: Physiology and Functional Morphology*, ed. B. Gartner (San Diego: Academic Press), 223–240. doi: 10.1016/b978-012276460-8/50012-6
- Pittermann, J., Limm, E., Rico, C., and Christman, M. A. (2011). Structure-function constraints of tracheid-based xylem: a comparison of conifers and ferns. *New Phytol.* 192, 449–461. doi: 10.1111/j.1469-8137.2011.03817.x
- Pryer, K. M., Schneider, H., Smith, A. R., Cranfill, R., Wolf, P. G., Hunt, J. S., et al. (2001). Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. *Nature* 409, 618–622. doi: 10.1038/35054555
- Renzaglia, K. S., Johnson, T. H., Gates, H. D., and Whittier, D. P. (2001). Architecture of the sperm cell of *Psilotum*. *Am. J. Bot.* 88, 1151–1163. doi: 10.2307/3558326
- Rothwell, G. W., Millay, M. A., and Stockey, R. A. (2018). Resolving the overall pattern of marattiale fern phylogeny. *Am. J. Bot.* 105, 1304–1314. doi: 10.1002/ajb2.1115
- Shamsuddin, T., Khan, S. A., Ahmad, I., Rahman, W., and Shamsuddin, K. M. (1985). Psilotic acid, a C6-C4-acid from *Psilotum nudum*. *Phytochemistry* 24, 2458–2459. doi: 10.1016/S0031-9422(00)83070-8
- Siegel, S. M. (1976). Inhibitory activity of the arylpyrone glucoside psilotin and its reversal by gibberellic acid and thiols. *Phytochemistry* 15, 566–567. doi: 10.1016/S0031-9422(00)88980-3
- Takahashi, M., Yamane, H., Satoh, Y., Takahashi, N., and Iwatsuki, K. (1984). Identification of GA36 in *Psilotum nudum*. *Phytochemistry* 23:681. doi: 10.1016/S0031-9422(00)80407-0
- Takahashi, S., Nakamura, F., Sahashi, N., Ohmoto, T., Mizushima, U., Sankawa, U., et al. (1990). Chemical markers of the Psilotaceae. *Biochem. Syst. Ecol.* 18, 11–12. doi: 10.1016/0305-1978(90)90024-A
- Tse, A., and Towers, G. H. N. (1967). The occurrence of psilotin in *Tmesipteris*. *Phytochemistry* 6:149. doi: 10.1016/0031-9422(67)85023-4
- Wallace, J. W., and Markham, K. R. (1978). Apigenin and amentoflavone glycosides in the psilotaceae and their phylogenetic significance. *Phytochemistry* 17, 1313–1317. doi: 10.1016/S0031-9422(00)94580-1
- Wang, G., Yao, S., Zhang, X. X., and Song, H. (2015). Rapid screening and structural characterization of antioxidants from the extract of *Selaginella doederleinii* Hieron with DPPH-UPLC-Q-TOF/MS method. *Int. J. Anal. Chem.* 2015:849769. doi: 10.1155/2015/849769
- Waterman, M. J., Nugraha, A. S., Hendra, R., Ball, G. E., Robinson, S. A., and Keller, P. A. (2017). Antarctic moss biflavonoids show high antioxidant and ultraviolet-screening activity. *J. Nat. Prod.* 80, 2224–2231. doi: 10.1021/acs.jnatprod.7b00085
- White, R. A., and Turner, M. D. (1995). Anatomy and development of the fern sporophyte. *Bot. Rev.* 61, 281–305. doi: 10.1007/BF02912620
- Wink, M. (1993). The plant vacuole: a multifunctional compartment. *J. Exp. Bot.* 44, 231–246.
- Yamaguchi, L. F., Kato, M. J., and Di Mascio, P. (2009). Biflavonoids from *Araucaria angustifolia* protect against DNA UV-induced damage. *Phytochemistry* 70, 615–620. doi: 10.1016/j.phytochem.2009.03.003
- Yu, S., Yan, H., Zhang, L., Shan, M., Chen, P., Ding, A., et al. (2017). A review on the phytochemistry, pharmacology, and pharmacokinetics of amentoflavone, a naturally-occurring biflavonoid. *Molecules* 22:299. doi: 10.3390/molecules22020299
- Yu, X., Wang, J., Zhang, J., Wang, L., Wang, Z., and Xiong, F. (2015). Physicochemical properties of starch isolated from bracken (*Pteridium aquilinum*) rhizome. *J. Food Sci.* 80, C2717–C2724. doi: 10.1111/1750-3841.13129
- Zhang, Y. X., Li, Q. Y., Yan, L. L., and Shi, Y. (2011). Structural characterization and identification of biflavones in *Selaginella tamariscina* by liquid chromatography-diode-array detection/electrospray ionization tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* 25, 2173–2186. doi: 10.1002/rcm.5090
- Zhang, S., Zhong, G., Liu, B., and Wang, B. (2011). Physicochemical and functional properties of fern rhizome (*Pteridium aquilinum*) starch. *Starch/Stärke* 63, 468–474. doi: 10.1002/star.201000142

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Šamec, Pierz, Srividya, Wüst and Lange. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Formation of Flavonoid Metabolons: Functional Significance of Protein-Protein Interactions and Impact on Flavonoid Chemodiversity

Toru Nakayama*, Seiji Takahashi and Toshiyuki Waki

Department of Biomolecular Engineering, Graduate School of Engineering, Tohoku University, Sendai, Japan

OPEN ACCESS

Edited by:

Kazuki Saito,
RIKEN Center for Sustainable
Resource Science (CSRS), Japan

Reviewed by:

Taira Miyahara,
Chiba University, Japan
Stefan Martens,
Fondazione Edmund Mach, Italy

*Correspondence:

Toru Nakayama
toru.nakayama.e5@tohoku.ac.jp

Specialty section:

This article was submitted to
Plant Metabolism and
Chemodiversity,
a section of the journal
Frontiers in Plant Science

Received: 19 March 2019

Accepted: 07 June 2019

Published: 09 July 2019

Citation:

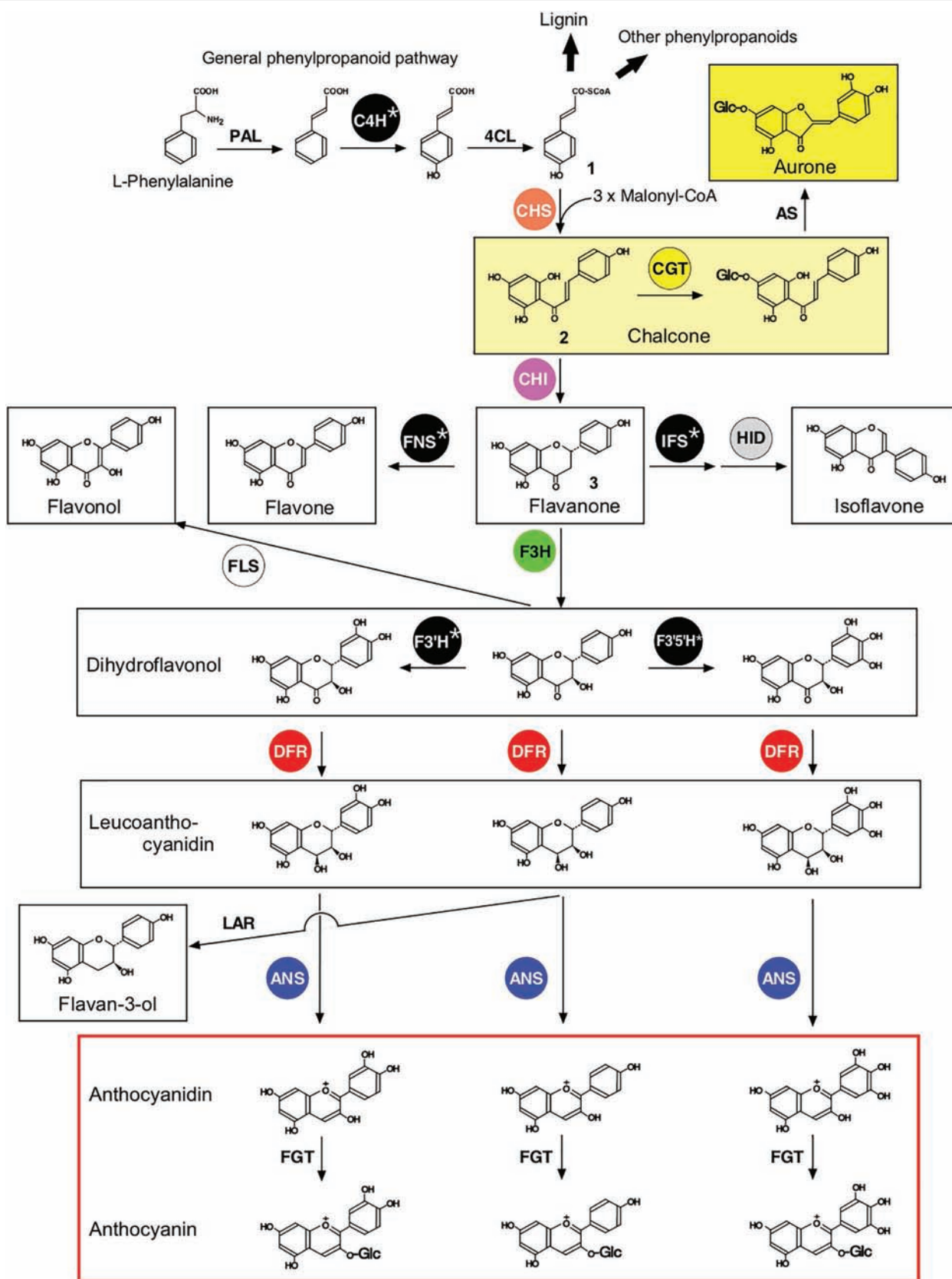
Nakayama T, Takahashi S and Waki T
(2019) Formation of Flavonoid
Metabolons: Functional Significance
of Protein-Protein Interactions and
Impact on Flavonoid Chemodiversity.
Front. Plant Sci. 10:821.
doi: 10.3389/fpls.2019.00821

Flavonoids are a class of plant specialized metabolites with more than 6,900 known structures and play important roles in plant survival and reproduction. These metabolites are derived from *p*-coumaroyl-CoA via the sequential actions of a variety of flavonoid enzymes, which have been proposed to form weakly bound, ordered protein complexes termed flavonoid metabolons. This review discusses the impacts of the formation of flavonoid metabolons on the chemodiversity of flavonoids. Specific protein-protein interactions in the metabolons of *Arabidopsis thaliana* and other plant species have been studied for two decades. In many cases, metabolons are associated with the ER membrane, with ER-bound cytochromes P450 hypothesized to serve as nuclei for metabolon formation. Indeed, cytochromes P450 have been found to be components of flavonoid metabolons in rice, snapdragon, torenia, and soybean. Recent studies illustrate the importance of specific interactions for the efficient production and temporal/spatial distribution of flavonoids. For example, in diverse plant species, catalytically inactive type-IV chalcone isomerase-like protein serves as an enhancer of flavonoid production via its involvement in flavonoid metabolons. In soybean roots, a specific isozyme of chalcone reductase (CHR) interacts with 2-hydroxyisoflavanone synthase, to which chalcone synthase (CHS) can also bind, providing a mechanism to prevent the loss of the unstable CHR substrate during its transfer from CHS to CHR. Thus, diversification in chemical structures and temporal/spatial distribution patterns of flavonoids in plants is likely to be mediated by the formation of specific flavonoid metabolons via specific protein-protein interactions.

Keywords: metabolon, flavonoids, chemodiversity, biosynthesis, protein-protein interaction, binary interaction, cytochrome P450, ER

INTRODUCTION

Flavonoids are a class of plant specialized metabolites with a basic C6-C3-C6 skeleton, for which 10 major classes (i.e., chalcones, aurones, flavanones, flavones, isoflavones, dihydroflavonols, flavonols, leucoanthocyanidins, anthocyanidins, and flavan-3-ols) have been described (**Figure 1**). In nature, flavonoids generally occur as glycosides or acylglycosides, with more than 6,900



(Continued)

FIGURE 1 | Proposed general pathways of flavonoid biosynthesis. Ten flavonoid classes are shown within boxes. Enzymes that are discussed in terms of their protein-protein interactions in this review are shown within circles. Enzyme abbreviations are: PAL, phenylalanine ammonia lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate:CoA ligase; CHS, chalcone synthase; CGT, chalcone 4'-O-glucosyltransferase; AS, aureusidin synthase; IFS, 2-hydroxyisoflavanone synthase; HID, 2-hydroxyisoflavanone dehydratase; CHI, chalcone isomerase; FNS, flavone synthase; F3H, flavanone 3-hydroxylase; FLS, flavonol synthase; F3'H, flavonoid 3'-hydroxylase; F3'5'H, flavonoid 3',5'-hydroxylase; DFR, dihydroflavonol 4-reductase; ANS, anthocyanidin synthase; LAR, leucoanthocyanidin 4-reductase; FGT, flavonoid 3-O-glucosyltransferase. Note that F3'H and F3'5'H may act on flavanones, flavones, and flavonols, depending on the plant species (not shown). Flavonoids and related metabolites are *p*-coumaroyl-CoA (**1**), 2',4,4',6'-tetrahydroxychalcone (THC) (**2**), and naringenin (**3**).

different structures (Arita and Suwa, 2008). Each plant lineage produces structurally distinct flavonoids in a lineage-specific manner, which play important roles in plant survival and reproduction. For example, in many cases, flower colors arise from anthocyanins and other flavonoids, which contribute to attracting pollinators (Andersen and Markham, 2006). In legumes, (iso)flavonoids in root exudates serve as chemoattractants for specific symbiotic bacteria as well as genetic inducers of nodulation (Barz and Welle, 1992; Subramanian et al., 2006, 2007). These (iso)flavonoids also play important roles in plant defensive mechanisms against infections by pathogens and invasion by herbivores (Aoki et al., 2000). Moreover, consumption of flavonoids is relevant for human nutrition, as illustrated by soybean [*Glycine max* (L.) Merr.] isoflavones, which exhibit estrogen-like and antioxidant activities and have been implicated in the ability of soy to prevent hormone-dependent cancers and cardiovascular diseases (Wiseman, 2006). These diverse bioactivities of flavonoids in plant biology and human nutrition are closely related to their diversity in chemical structure.

Flavonoids are derived from the amino acid L-phenylalanine via the general phenylpropanoid pathway, shown in **Figure 1** (Winkel-Shirley, 2001). Chalcone synthase (CHS), the first committed enzyme of the flavonoid pathway, catalyzes the production of 2',4,4',6'-tetrahydroxychalcone (THC, **2**; **Figure 1**), which serves as a precursor for the other flavonoids (Austin and Noel, 2003). Aurones are directly derived from chalcones in limited plant species (Nakayama et al., 2000, 2001; Kaintz et al., 2014), while other flavonoids, including flavones, isoflavones, flavonols, and anthocyanidins, are derived after the conversion of chalcones to flavanones catalyzed by chalcone isomerase (CHI) (Winkel-Shirley, 2001). While the core flavonoid pathway is well conserved among seed plants, specific lineages develop specific flavonoid pathways to enhance fitness in particular environmental conditions. Enzymes involved in flavonoid biosynthesis (**Figure 1**) include polyketide synthases (e.g., CHS), 2-oxoglutarate-dependent dioxygenases [e.g., flavanone 3-hydroxylase (F3H, also termed FHT), anthocyanidin synthase (ANS; also termed leucoanthocyanidin dioxygenase, LDOX), flavonol synthase (FLS), flavone synthase I (FNSI)], short-chain dehydrogenases/reductases [e.g., dihydroflavonol 4-reductase (DFR)], aldo-keto reductases [e.g., chalcone reductase (CHR)], and cytochrome P450 monooxygenases [e.g., flavone synthase II (FNSII), flavonoid 3'-hydroxylase (F3'H), flavonoid 3',5'-hydroxylase (F3'5'H), and 2-hydroxyisoflavanone synthase (IFS)]. These enzymes are hypothesized to have evolved from enzymes involved in primary metabolism (Weng and Noel, 2012; Moghe and Last, 2015). Cytochromes P450, shown with an asterisk in **Figure 1**, have been shown to be anchored to the cytoplasmic surface of the ER (Ralston and Yu, 2006),

while most of the other enzymes are proposed to be soluble enzymes. A variety of regio-specific glycosyltransferases, acyltransferases, methyltransferases, and prenyltransferases acting on flavonoids have evolved in a lineage-specific manner to enhance the structural diversity of flavonoids (Ono et al., 2010; Sasaki and Nakayama, 2015).

It is generally accepted that the intracellular environments are of macromolecular crowding state (Fulton, 1982). Given our understanding of diffusion rates of small solutes and macromolecules in cells and organelles (Verkman, 2002), it is now recognized that cells and organelles are not simply bags of enzymes; rather, metabolic enzymes in the same pathway tend to be associated with each other in cellular environments, with each of these metabolic pathways confined to a specific region of the cell (microcompartmentalization of cellular metabolism) (Saks et al., 2008). The weakly bound, ordered complexes of enzymes involved in sequential metabolic pathways are referred to as “metabolons” (Ovadi and Srere, 2000; Srere, 2000; Ovadi and Saks, 2004; Jørgensen et al., 2005; Sweetlove and Fernie, 2013). The formation of a metabolon is believed to provide catalytic advantages via substrate channeling, including preventing the loss of intermediates by diffusion, reducing the transit time between active sites, protecting the chemically labile intermediates, circumventing unfavorable equilibria, and segregating the intermediates of competing reactions (Ovadi, 1991). The formation of metabolons is well defined in primary metabolic pathways of prokaryotic and eukaryotic cells, including glycolysis (Giege et al., 2003; Graham et al., 2007), the tricarboxylic acid cycle (Wu et al., 2015; Wu and Menteer, 2015), the Calvin-Benson cycle (Suss et al., 1993), and nucleotide synthesis (An et al., 2010). In plant specialized metabolism, metabolons formed during the biosynthesis of cyanogenic glycosides (Laursen et al., 2016) and lignins (Gou et al., 2018) in *Arabidopsis thaliana* (L.) Heynh. and other plant species have been studied in detail (Sweetlove and Fernie, 2013). In many cases, metabolon formation takes place on biological membranes or cytoskeletal elements via specific interactions of soluble enzymes with these cellular structures. However, because protein-protein interactions in metabolons are weak in most cases, it is difficult to isolate metabolons in their intact forms.

The concept of flavonoid metabolons was first proposed in 1974 to explain the efficiency of flavonoid synthesis in plant cells (Stafford, 1974). Subsequently, the association of flavonoid enzymes on biological membranes (e.g., the ER) and the formation of complexes were supported by several lines of experimental evidence (Hrazdina and Wagner, 1985a,b; Hrazdina et al., 1987; reviewed by Winkel, 2004). Since then, flavonoid metabolons have been assumed to form in diverse plant species;

a model of flavonoid metabolon was proposed as a linear array of consecutive flavonoid enzymes along the ER (Hrazdina and Wagner, 1985a,b; Stafford, 1990). To date, specific protein-protein interactions in flavonoid metabolons have been studied in multiple plant species. Substrate channeling between DFR and leucoanthocyanidin 4-reductase (LAR) was predicted by computational studies, which also suggested the functional significance of metabolon formation during flavonoid synthesis (Diharce et al., 2016). Thus, elucidation of the structural organization of metabolons provides a basis for understanding how flavonoid structures are diversified, as well as how the temporal and spatial accumulations of flavonoids are regulated (Laursen et al., 2015). This review describes our current knowledge of specific protein-protein interactions identified in flavonoid metabolons and discusses their functional significance in flavonoid biosynthesis.

CYTOCHROMES P450 CAN BE COMPONENTS OF FLAVONOID METABOLONS

It has been shown so far that soluble enzymes involved in plant specialized metabolism are associated on the cytoplasmic surface of ER to form metabolons, nucleated by ER-bound cytochromes P450. More than three decades ago, some of the soluble enzymes related to the general phenylpropanoid and flavonoid pathways, L-phenylalanine ammonia-lyase (PAL), CHS, and flavonoid glucosyltransferase, were found to be associated with the ER membrane in several plant species including *Hippeastrum* (amaryllis, order Asparagales) and *Fagopyrum esculentum* (order Caryophyllales) (Hrazdina and Wagner, 1985a,b; Hrazdina et al., 1987), suggesting the occurrence of ER-bound metabolons for the synthesis of phenylpropanoids and flavonoids. Meanwhile, the formation of metabolons during the syntheses of other classes of plant specialized metabolites, including cyanogenic glucosides and lignins, was shown to involve the anchoring of soluble enzymes by cytochromes P450 to specific domains of the ER membrane [reviewed by Ralston and Yu (2006)]. In 2004, in tobacco (*Nicotiana tabacum*, order Solanales), cinnamate 4-hydroxylase (C4H), a cytochrome P450 (CYP73A) that is involved in the general phenylpropanoid pathway (Figure 1), was found to be responsible for the weak association of soluble isozymes of PAL (PAL1 and PAL2) with ER membranes (Figure 1), using a combination of biochemical and fluorescence microscopic methods (Achnine et al., 2004).

Formation of a flavonoid metabolon on cytochrome P450 was first demonstrated in 2008 in rice (*Oryza sativa* L.; order Poales, a monocot) that accumulates flavones, flavonols, proanthocyanidins (oligomeric flavan-3-ols), and anthocyanins (Shih et al., 2008). In this plant, an isozyme of flavonoid 3'-hydroxylase (F3'H1), a cytochrome P450 (CYP75B) catalyzing the 3'-hydroxylation of the B-ring of flavonoids (Figure 1), was shown to bind to CHS1 (a CHS isozyme) (Figure 2A) by yeast two-hybrid (Y2H) assays (Shih et al., 2008). The series of binary interaction assays showed that CHS1 also interacts

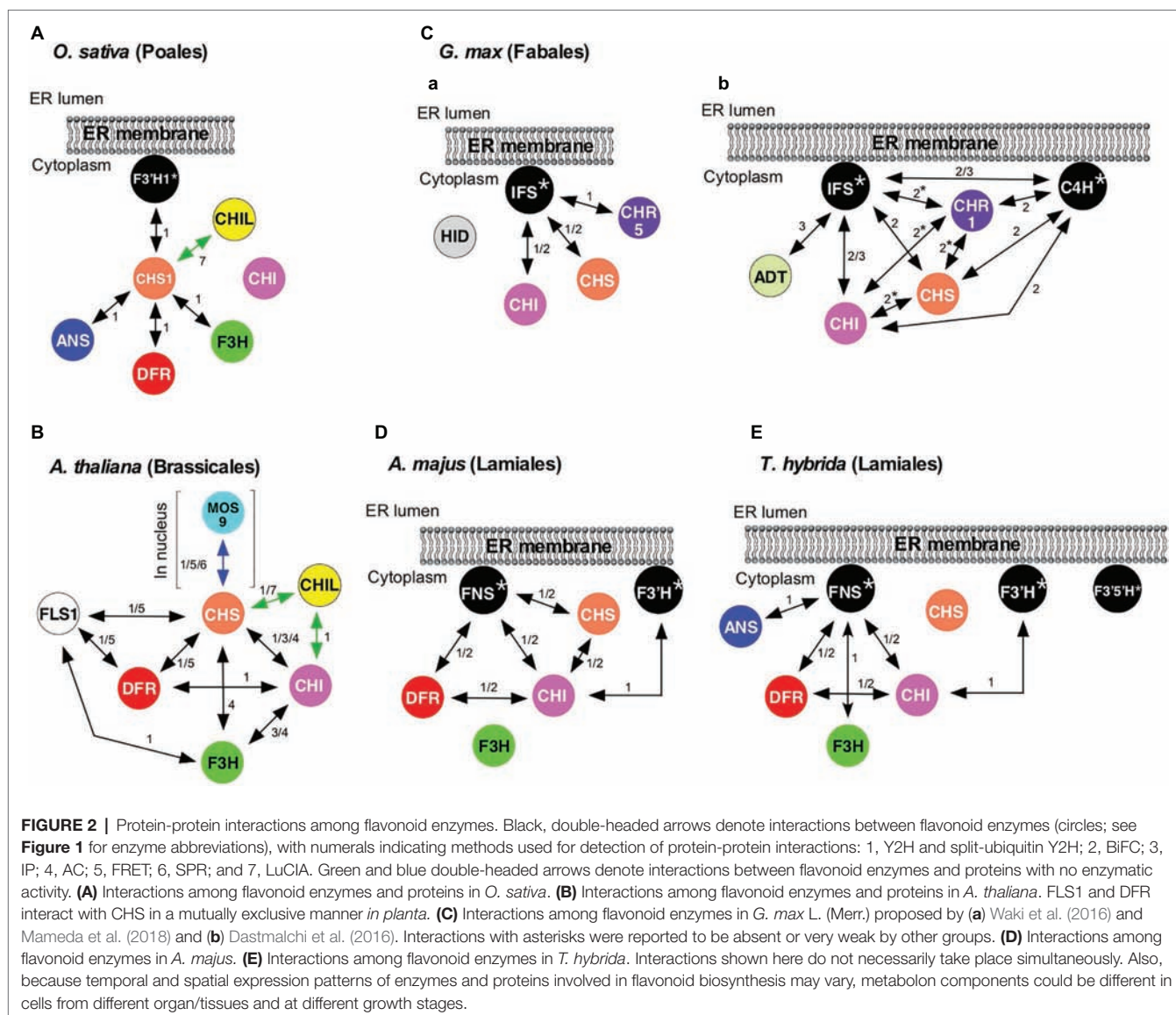
with F3H, DFR, and ANS1 (an isozyme of rice ANS), but not with CHI (Figure 2A). Interactions among CHI, F3H, F3'H1, DFR, and ANS1 were not detected. It was proposed that in rice, CHS could serve as a common platform for a flavonoid metabolon, which might be anchored to the cytoplasmic surface of the ER via F3'H1. In 2016, two groups independently published evidence supporting the association of soybean flavonoid enzymes in metabolons tethered to the ER-bound cytochromes P450 IFS (CYP93C) and C4H (Figures 2Ca,Cb; Dastmalchi et al., 2016; Waki et al., 2016; Mamedea et al., 2018). Additionally, physical interactions among flavonoid enzymes in snapdragon (*Antirrhinum majus* L.) and torenia (*Torenia hybrida*) were clarified, in which FNSII (CYP93B1, Figures 2D,E) was found to be a component of flavonoid metabolons (Fujino et al., 2018) (see below).

FLAVONOID DIVERSITY AND FLAVONOID METABOLONS

During the past two decades, physical interaction partnerships of flavonoid enzymes and related proteins have been studied in multiple phylogenetically distinct plants, including rice, *A. thaliana*, soybean, snapdragon, and hops (*Humulus lupulus* L. var. *lupulus*), each of which belongs to different orders of plants and accumulates different classes of flavonoids (see below). The data suggest that production of specific flavonoids in these plants is attained via spatially and temporally dependent interactions between specific proteins during plant growth and stress responses. These data are discussed in more detail below.

FLAVONOID METABOLONS IN ARABIDOPSIS

A. thaliana (order Brassicales), in which 54 flavonoid species have been identified to date, primarily accumulates flavonols and proanthocyanidins and also produces anthocyanins under stress conditions (Saito et al., 2013). Studies of *A. thaliana* flavonoid enzymes provide one of the best-characterized flavonoid metabolons with respect to protein-protein interactions. Direct protein-protein interactions among soluble flavonoid enzymes in *A. thaliana* have been studied by Y2H assays, affinity chromatography (AC), immunoprecipitation (IP), and physicochemical methods: Förster resonance energy transfer (FRET) detected by fluorescence lifetime imaging microscopy (FLIM) and surface plasmon resonance refractometry (SPR) (Burbulis and Winkel-Shirley, 1999; Owens et al., 2008; Crosby et al., 2011; Watkinson et al., 2018). In this plant, interactions between the following enzyme pairs have been identified (followed by methods in parentheses): CHS-CHI (Y2H, AC, and IP), CHS-F3H (AC), CHS-DFR (Y2H and FRET), CHS-isozyme of FLS (FLS1) (Y2H and FRET), CHI-DFR (Y2H), CHI-F3H (AC and IP), FLS1-DFR (Y2H and FRET), and FLS1-F3H (Y2H) (Figure 2B, black arrows). Interactions of catalytically inactive paralogs of FLS with CHS and DFR were



also found *via* Y2H (Owens et al., 2008). These binary interactions suggested a flavonoid metabolon model (**Figure 2B**) with CHS as the hub. Moreover, this model features a globular association, rather than a linear array, of flavonoid enzymes in the metabolons. FRET-FLIM analyses revealed that FLS1 and DFR, the key enzymes of branch pathways (**Figure 1**), interact with CHS in a mutually exclusive manner *in planta* (Crosby et al., 2011). This provides a possible *in planta* mechanism for regulating metabolic flux by changing physical interactors with CHS, which is pivotal in the pathway.

It has been shown that not only flavonoid enzymes but also a protein with no catalytic activity can be a component of the flavonoid metabolon in *A. thaliana*. Phylogenetic analyses suggest that CHI enzymes have evolved from a non-catalytic ancestor related to fatty acid-binding proteins (FAPs) and land plant-specific CHI-like proteins (CHILs) with no catalytic activity (Ngaki et al., 2012; Kaltenbach et al., 2018). Thus, CHIs, FAPs,

and CHILs, all of which are soluble proteins, constitute a larger structurally related family, the CHI-fold family, in which CHIs correspond to types I and II, and FAPs and CHILs, respectively, correspond to types III and IV within the family. In 2014, CHILs were shown to serve as enhancers of flavonoid production (EFPs), as loss-of-function mutations and suppression in morning glory (*Ipomoea nil*) and torenia, respectively, resulted in a significant diminution of flavonoid contents (Morita et al., 2014). CHIL is also produced by *A. thaliana*. Y2H analyses indicated that in *A. thaliana*, CHIL binds to CHI (**Figure 2B**, green arrow), suggesting that CHIL is a component of the flavonoid metabolon (Jiang et al., 2015). Recently, CHIL of *A. thaliana* (AtCHIL) was also shown to physically interact with CHS of the same plant species (AtCHS) (**Figure 2B**, green arrow) by Y2H and luciferase-complementation imaging assays (LuCIA) (Ban et al., 2018). The coexpression of AtCHIL with AtCHS in yeast *Saccharomyces cerevisiae* resulted in a 1.8-fold enhancement

of AtCHS-catalyzed production of THC. The interactions of CHIL with CHI and CHS might be related to the observed role of CHIL as an EFP. It must be mentioned that the binding of CHIL to CHS has also been observed in the flavonoid systems of hops, rice (**Figure 2A**), *Selaginella moellendorffii* (a lycophyte), and *Physcomitrella patens* (a bryophyte), as assayed by LuCIA (Ban et al., 2018). The coexpression of CHILs with CHS of these plants in *S. cerevisiae* also enhanced the CHS-catalyzed production of THC, suggesting the conservation of the EFP role of CHIL proteins throughout land plants.

The dynamism and versatility of CHS-mediated protein-protein interactions likely take place in organ- and organelle-specific manners in *A. thaliana*. Immunofluorescence and immunoelectron microscopic analyses showed that CHS and CHI co-localize at the ER and tonoplasts in epidermal and cortex cells of *A. thaliana* roots (Saslawsky and Winkel-Shirley, 2001). This observation suggests that a subset of CHS and CHI enzymes in root cells might not be assembled into metabolons that are mentioned above. As both of these enzymes are soluble, these data suggest that one or more other proteins function in recruiting these enzymes to membranes, although this protein remains to be identified. It remains to be determined whether CHS and CHI interact with cytochromes P450 in *A. thaliana*. In this context, in *A. thaliana* root cells, F3'H is unlikely to be involved in recruiting CHS and CHI to the ER, as suggested by the results of immunolocalization in the *A. thaliana* F3'H mutant *tt7(88)* (Saslawsky and Winkel-Shirley, 2001).

CHS and CHI have also been shown in the nucleus of *A. thaliana* by multiple immunolocalization methods (Saslawsky et al., 2005). CHS of *A. thaliana*, a dimeric enzyme, possesses sequences resembling a nuclear localization signal, which is located on the surface opposite from the dimerization interface. This signal could direct CHS and associated enzymes into the nucleus. Moreover, immunoblotting of nuclear CHI suggested post-translational modifications that also might be responsible for the nuclear localization of the enzyme. Interestingly, CHS was recently found to interact with MOS9 (a nuclear protein associated with epigenetic control of *R* genes that mediate effector-triggered immunity) as analyzed by Y2H, SPR, and FRET, with a K_d of 210 nM (**Figure 2B**, blue arrow) (Watkinson et al., 2018). Addressing this finding further may uncover additional mechanisms for controlling flavonoid pathways, as well as linking them to defense mechanisms and other physiological functions.

THE SOYBEAN ISOFLAVONOID METABOLON

The soybean (order Fabales) produces isoflavones, which are a class of flavonoids with a 3-phenylchromone structure and distributed almost exclusively in legumes (Aoki et al., 2000). Isoflavones play important roles in symbiotic plant-microbe interactions and defensive mechanisms against pathogen infection in soybean (Barz and Welle, 1992). Moreover, soybean isoflavones show a variety of bioactivities that are beneficial

to human health (Wiseman, 2006). The soybean produces two distinct types of isoflavonoids: 5-deoxyisoflavonoids (daidzein and its conjugates) and 5-hydroxyisoflavonoids (genistein and its conjugates) (**4** and **8**, respectively, **Figure 3A**). In unstressed soybean plants (cv. Enrei), 5-deoxyisoflavonoids accumulate in the roots (93% mol/mol of total root isoflavonoids) and seeds (60% mol/mol of the total seed isoflavonoids) (Mameda et al., 2018).

Characterization of Isoflavonoid Metabolon

Protein-protein interaction analyses of soybean isoflavonoid enzymes suggested that biosynthesis of isoflavones takes place *via* the formation of a metabolon on cytochromes P450. Specifically, the analysis using split-ubiquitin Y2H and bimolecular fluorescence complementation (BiFC) assay systems revealed that each enzyme located upstream of the isoflavonoid pathway (CHS, CHI, and GmCHR5 (an isozyme of soybean CHR); **Figures 1, 3A**) interacts with isozymes of IFS (CYP93C) to form a metabolon (**Figure 2Ca**; Waki et al., 2016; Mameda et al., 2018). It has been proposed that C4H also serves as a nucleus for the metabolon formation as analyzed *via* BiFC and IP (**Figure 2Cb**; Dastmalchi et al., 2016). Moreover, arogenate dehydratase (ADT), a shikimate pathway enzyme that has primarily been reported to be a plastidial enzyme, was reported to interact with IFS on the basis of IP (**Figure 2Cb**; Dastmalchi et al., 2016). The fluorescence localizations observed during these BiFC analyses were consistent with P450-mediated interactions taking place at the ER (Waki et al., 2016). As the activities of IFS and C4H are indispensable for the formation of isoflavones, these cytochromes P450 are considered to play both catalytic and structural roles in the metabolon.

The affinity of the soybean isoflavonoid enzymes for IFS isozymes varies among paralogs. Isoflavonoid enzymes shown in **Figure 2C** are encoded by multiple genes in soybean. For example, there are at least nine paralogs (GmCHSs) encoding CHS (Schmutz et al., 2010; Shimomura et al., 2015), 12 encoding CHI (Shimada et al., 2003; Ralston et al., 2005), and two encoding IFS (Cheng et al., 2008). For each enzyme, different isozymes exert different physiological functions (Shimizu et al., 1999; Tuteja et al., 2004; Livingstone et al., 2010). The analysis using split-ubiquitin Y2H system suggested GmCHS1 has a higher affinity for GmIFS1 than for GmCHS7 (Waki et al., 2016), and GmCHR5 binds to GmIFS isozymes but other GmCHR isozymes cannot (see below for details) (Mameda et al., 2018). These observations could be related to differential regulation and physiological roles of each enzyme paralog.

An Implication for Functional Significance of Protein-Protein Interactions During 5-Deoxyisoflavonoid Biosynthesis

Although 5-deoxyisoflavonoids accumulate in the roots and seeds of unstressed plants in a high ratio (Mameda et al., 2018), its mechanistic details remained unknown. During the course of 5-deoxyisoflavonoid biosynthesis, isoliquiritigenin (**Figure 3A, 6**) (a 6'-deoxychalcone) is produced *via* a

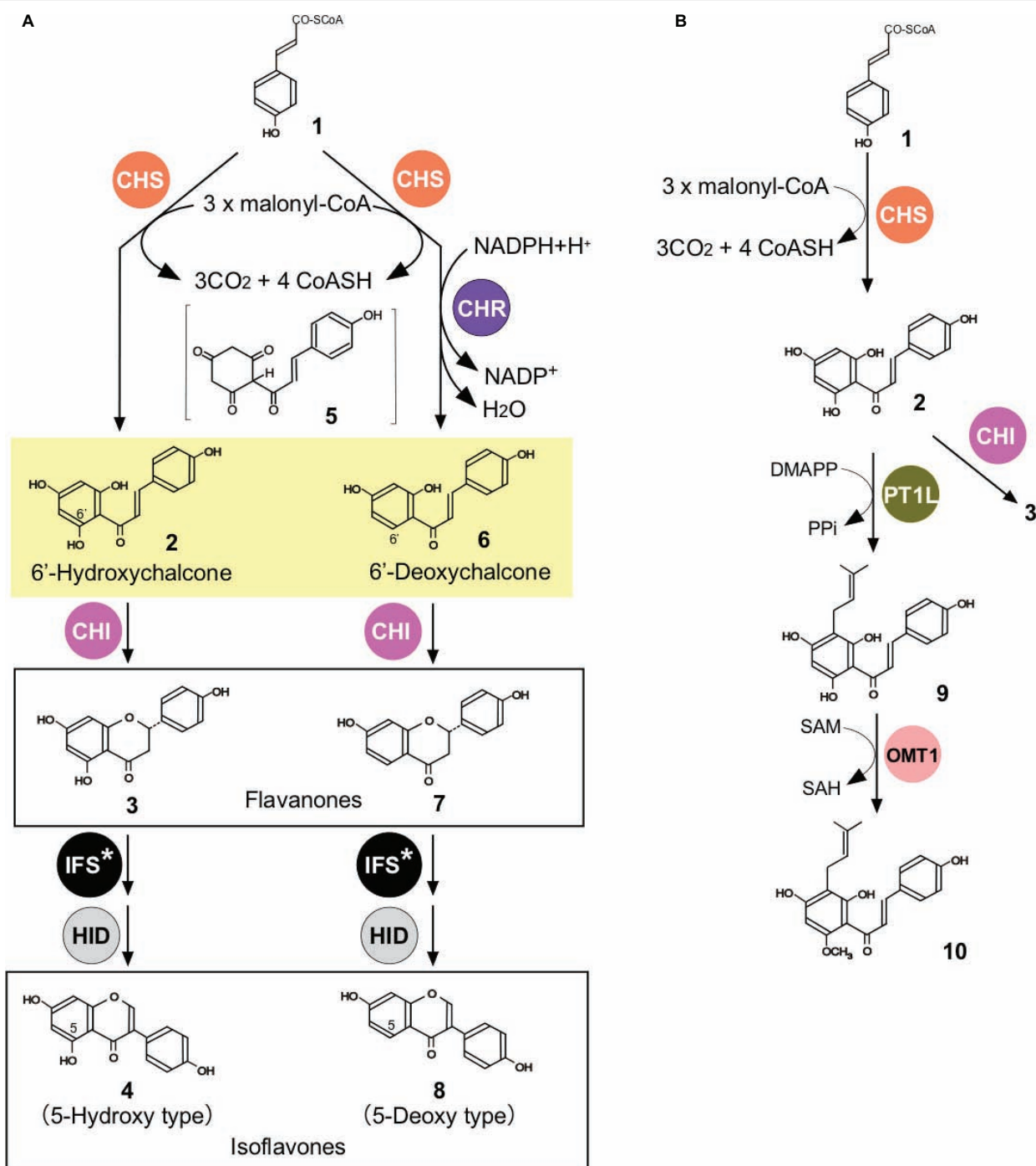


FIGURE 3 | Biosynthesis of flavonoids in soybean **(A)** and hops **(B)**. **(A)** Biosyntheses of 5-hydroxy- and 5-deoxyisoflavonoids in soybean. CHR, chalcone reductase. See **Figure 1** for abbreviations for other enzymes. **(B)** Biosyntheses of prenylated flavonoids in hops. PT, aromatic prenyltransferase; OMT, O-methyltransferase; DMAPP, dimethylallyl diphosphate; PPI, pyrophosphate; SAM, S-adenosyl-L-methionine; SAH, S-adenosyl-L-homocysteine. Flavonoids and related metabolites are *p*-coumaroyl-CoA (1), THC (2), naringenin (3), genistein (4), *p*-coumaroylcyclohexantrione (5), isoliquiritigenin (6), liquiritigenin (7), daidzein (8), demethylxanthohumol (9), and xanthohumol (10).

CHS-catalyzed reaction coupled to CHR catalysis (Bomati et al., 2005). The soybean genome encodes 11 CHR paralogs (Mameda et al., 2018), among which only GmCHR1 had been characterized enzymatically (Welle and Grisebach, 1988; Welle et al., 1991).

Although CHR has been referred to as chalcone reductase, it does not actually act on THC (2, **Figure 3A**) but instead on one of the diffusible intermediates of the CHS-catalyzed reaction, most likely *p*-coumaroylcyclohexantrione (5, **Figure 3A**),

which is highly unstable and is rapidly aromatized to produce THC in an aqueous system (Bomati et al., 2005). THC and isoliquiritigenin then undergo the reactions catalyzed by CHI, IFS, and 2-hydroxyisoflavanone dehydratase (HID) to produce genistein (4) and daidzein (8), respectively (Figure 3A; Mameda et al., 2018). The amount of the CHR product isoliquiritigenin does not generally exceed 25% (mol/mol) of the total CHS products (isoliquiritigenin, THC, and naringenin) during the combined action of dilute CHS and GmCHR1 (0.05 μ M each) *in vitro*. These low product ratios for CHR catalysis during *in vitro* assays could arise from the fact that only a small fraction (<25%) of 5 produced during CHS catalysis is transferred to the active site of GmCHR1 while the majority (>75%) escapes and diffuses to the aqueous system to give rise to THC.

To establish a high 5-deoxyisoflavonoid ratio in the cells of soybean roots and seeds (i.e., a high product ratio for CHR catalysis), 5 has to be immediately transferred, prior to aromatization, from the active site of CHS to that of CHR. One possible mechanism for achieving this would be binding of CHR to CHS, facilitating the channeling of 5 between them. However, the crystal structure of CHR suggested that direct association of the active sites of CHR and CHS is impossible and that passive diffusion may be the only way to transfer 5 from CHS to CHR (Bomati et al., 2005). In fact, Y2H assays showed that GmCHR1 (the only GmCHR paralog whose catalytic activity was confirmed) was unable to interact with any of the GmCHS isozymes (Waki et al., 2016; Mameda et al., 2018). Alternatively, a shorter distance or transit time for 5 between the two enzymes could be achieved in a metabolon and located very close to each other. Because CHS isozymes have been shown to interact with IFS isozymes (GmIFS) (Figure 2Ca; Dastmalchi et al., 2016; Waki et al., 2016), the involvement of GmCHR1 in the isoflavonoid metabolon was examined. However, GmCHR1 was not found to interact with any of the enzymes examined including IFS isozymes (Waki et al., 2016; Mameda et al., 2018). Moreover, the product ratio for CHR catalysis did not exceed 50% even when high concentrations of GmCHR1 and CHS were used in *in vitro* enzyme assays (Oguro et al., 2004; Mameda et al., 2018). Therefore, the involvement of GmCHR1 in the observed high proportion of 5-deoxyisoflavonoids in the roots and seeds of unstressed plants was unlikely.

Thus, 11 GmCHR paralogs were comprehensively analyzed for their possible involvement in biosynthesis of 5-deoxyisoflavonoids in the roots and seeds of unstressed plants, and the data obtained strongly suggested the involvement of a previously unappreciated soybean CHR, GmCHR5 (Figure 2Ca). Specifically, among the GmCHR paralogs examined, the expression patterns of *GmCHR5* were the most consistent with the observed patterns of the accumulation of daidzein conjugates in the roots and the seeds of unstressed plants. When interactions of these GmCHR isozymes with soybean isoflavonoid enzymes (Figure 3A) were analyzed by split-ubiquitin Y2H assays, GmCHR5 uniquely interacted with IFS isozymes (Mameda et al., 2018). Moreover, *in vitro* assay results suggested that the product ratio for CHR catalysis depended on the GmCHR5

concentration, with higher concentrations resulting in higher ratios (approaching 90%) (Mameda et al., 2018). Thus, the results of enzyme assays, transcription analyses, and protein-protein interaction assays were all consistent with the fact that GmCHR5, but not other CHR isozymes, is the key player in the accumulation of 5-deoxyisoflavonoids in the roots and seeds of unstressed plants. It would be highly likely that the interactions of CHS and GmCHR5 with IFS could allow the microcompartmentalization of the metabolic process, resulting in a product ratio for CHR catalysis high enough for the dominated accumulation of 5-deoxyisoflavonoids in the roots and seeds of unstressed plants. This illustrates the previously proposed functional significance of metabolon formation, i.e., preventing the loss of intermediates by diffusion and reducing the transit time between active sites. This also supports the hypothesis that specific spatial distributions of a flavonoid can be attained by inclusion of a specific isozyme in a flavonoid metabolon in a spatially specific manner.

Functional Differentiation of GmCHR Isozymes in the Soybean

GmCHR1 and GmCHR6 are unable to interact with any of the isoflavonoid enzymes shown in Figure 3A (Mameda et al., 2018). However, this does not necessarily rule out their involvement in 5'-deoxyisoflavonoid biosynthesis but is rather consistent with functional differentiation of GmCHR isozymes in the soybean. Previously, expression of *GmCHR1*, *GmCHR5*, and *GmCHR6* was shown to be induced upon microbial infection (Sepiol et al., 2017). Moreover, *GmCHR6* is located near a quantitative trait locus region linked to resistance to a pathogenic oomycete (Sepiol et al., 2017). In soybean, the production of both types (5-deoxy- and 5-hydroxy-) of isoflavonoids is induced by microbial pathogens. The production of both types of isoflavonoids, rather than the exclusive production of 5'-deoxy type, would be needed to fully implement relevant soybean defense mechanisms. The induced production of GmCHR1 and GmCHR6, showing the maximum product ratio for CHR catalysis of 50%, make it possible to accumulate high levels of both types of isoflavonoids in infected plants. Thus, it would be likely that GmCHR5 plays a key role in the exclusive accumulation of 5-deoxyisoflavonoids in the roots and seeds of unstressed plants while GmCHR1 and GmCHR6 play key roles in the induced defense mechanisms against microbial pathogens.

FLAVONOID METABOLONS IN THE ORDER LAMIALES

Snapdragon and torenia are flowering ornamentals in which colorful petals are the most eye-catching trait. The petal colors in these lamiales plants are mainly provided by flavonoids, which represent different flavonoid classes from those mainly found in *A. thaliana* and soybean. The petal colors of snapdragon—magenta, orange, red, pink, yellow,

cream, and white—are produced by a combination of anthocyanins (orange, pink, red, and reddish purple), aurones (yellow), and flavones (co-pigments) (Ono and Nakayama, 2007). *Torenia* accumulates anthocyanins and flavones in its flower petals, which are responsible for the bluish purple and pink colors (Ueyama et al., 2002).

In 2018, physical interactions among flavonoid enzymes in snapdragon and *torenia* were clarified, illustrating the formation of flavonoid metabolons responsible for flower coloration (Fujino et al., 2018). Binary interactions found in split-ubiquitin Y2H and BiFC assays were: FNSII-CHS, FNSII-CHI, FNSII-DFR, CHS-CHI, CHI-DFR, and F3'H-CHI in snapdragon; and FNSII-CHI, FNSII-F3H, FNSII-DFR, FNSII-ANS, CHI-DFR, and F3'H-CHI in *torenia* (Figures 2D,E). Split-ubiquitin Y2H assays also suggested that binding of CHI and DFR to FNSII is not exclusive in snapdragon.

Interestingly, enzymes involved in the late stage of anthocyanin biosynthesis (DFR in snapdragon; DFR, F3H, and ANS in *torenia*) were found to interact with FNSII (cytochrome P450 CYP93B1) (Figures 1, 2D,E; Fujino et al., 2018). The activity of FNSII is not needed for anthocyanin biosynthesis, suggesting that FNSII could function as a scaffold for anthocyanin biosynthesis. Although further studies are needed to test this hypothesis, several findings are consistent with FNSII as an important component of the metabolon for anthocyanin biosynthesis. Previously, attempts were made to engineer *torenia* flowers showing a deeper petal color using metabolic engineering (Ueyama et al., 2002). To achieve this, *FNSII* was co-suppressed in blue-violet *torenia* flowers to diminish FNSII activity. As anthocyanin synthesis competes with flavone synthesis for flavanones as the shared precursors (Figure 1), this was predicted to favor anthocyanin production at the expense of flavone formation (see Figure 1). This strategy was inspired by the observations in black dahlia (*Dahlia variabilis*, order Asterales) accumulating large amounts of anthocyanins, in which FNSII production is suppressed by endogenous posttranscriptional gene silencing (Thill et al., 2012; Deguchi et al., 2013). In the black dahlia, suppression of *FNSII* increased production of anthocyanins while flavone production was decreased. Metabolic engineering of *torenia* showed that the co-suppression of *FNSII* diminished flavone and increased flavanone levels in petals, as expected (Ueyama et al., 2002). However, anthocyanin levels in the petals of the *FNSII*-suppressed *torenia* decreased considerably, producing a paler flower. The reason for this result was unknown, but this observation can now be explained by FNSII acting as a component of the metabolon related to anthocyanin production.

Interestingly, in the anthocyanin-accumulating snapdragon petals, flavones were accumulated first, followed by anthocyanins, and finally aurones (Toki, 1988; Fujino et al., 2018). This sequence of flavonoid accumulation is consistent with the transcriptional patterns of snapdragon flavonoid enzyme genes during flower development (Fujino et al., 2018). Thus, on the basis of interactions (Figure 2D) and temporal gene expression patterns of flavonoid enzymes in red snapdragon petal cells, a model of the flower stage-dependent formation of the flavonoid metabolon has been proposed (Fujino et al., 2018). In this

model, CHS, CHI, and FNSII are expressed and form a flavone metabolon on the ER surface at the beginning of the flower development. Halfway through flower development, F3H and DFR are expressed to form an anthocyanin metabolon by using the preexisting flavone metabolon as a scaffold.

The similarity of interaction partnerships in the flavonoid metabolons of snapdragon and *torenia* (Figures 2D,E) is consistent with the close phylogenetic relationship of these plants. Collectively with the fact that the *A. thaliana* genome lacks *IFS* and *FNSII* genes, interactions in flavonoid metabolons may differ between plant species while those of closely related plant species are more similar to each other (Figure 2). This is consistent with the observed structural diversity of flavonoids in plants and the fact that each plant lineage produces structurally distinct flavonoids in a lineage-specific manner.

PROTEIN-PROTEIN INTERACTIONS OF FLAVONOID ENZYMES AND PROTEINS IN HOPS

Hops uniquely accumulate the prenylated flavonoids xanthohumol (3'-prenyl-6'-O-methyl-THC) (10, Figure 3B) and demethylxanthohumol (3'-prenyl-THC) (9) in the glandular trichomes (lupulin glands) of female cones, a key ingredient in beer brewing (Stevens and Page, 2004). Recent studies of the synthesis of these prenylated flavonoids provide examples of the involvement of non-catalytic CHI-fold proteins in flavonoid metabolons as specialized auxiliary proteins (Ban et al., 2018).

In hops, the soluble, trichome-specific isozyme of CHS (CHS_H1) is involved in the biosynthesis of prenylchalcones and catalyzes the production of THC (2, Figure 3B); THC is then prenylated by the membrane-bound, aromatic prenyltransferase PT1L, then 6'-O-methylated by the soluble O-methyltransferase OMT1 to produce xanthohumol (10, Figure 3B; Ban et al., 2018). Recent studies have shown that non-catalytic members of the CHI-fold protein family, CHIL1 (a type-III, FAP-related protein) and CHIL2 (a type-IV, EFP-related protein), are involved in the syntheses of prenylated flavonoids in hops. CHIL2 was found to interact with CHS_H1 and PT1L by Y2H, LuCIA, and IP assays (Ban et al., 2018). As PT1L is a membrane-bound enzyme with eight predicted transmembrane domains and proposed to localize in trichome plastids (Li et al., 2015), these results suggest a membrane-anchored metabolon for xanthohumol biosynthesis. *In vitro* enzymatic assays showed that CHIL2 slightly enhances the catalytic efficiencies of CHS_H1 and PT1L. Specifically, the binding of CHIL2 to CHS_H1 results in a 6–18-fold increase in k_{cat} and 5.5–6.0-fold increase in K_m for *p*-coumaroyl-CoA (1, Figure 3B) and malonyl-CoA, with up to 2.9-fold increase in k_{cat}/K_m values; whereas, the binding of CHIL2 to PT1L results in a slight increase in V_{max} and slight decrease in K_m for these substrates, with up to a 2.2-fold increase in V_{max}/K_m values. *S. cerevisiae* was engineered for the production of demethylxanthohumol. The engineered yeast co-expressing CHIL2 and CHS_H1 with PT1L produced greater amounts of demethylxanthohumol than those expressing CHS

alone, consistent with CHIL2 functioning as an EFP *in vivo*. Thus, specific binding of CHIL2 to CHS enhances the rate of CHS-catalyzed entry from the general phenylpropanoid pathway to the flavonoid pathway (Figure 1) to potentiate flavonoid production. Unlike CHIL2, CHIL1 did not interact with CHIL2, PT1L, CHS_H1, *p*-coumaroyl-CoA ligase (see Figure 1), or OMT, as found by multiple methods (Ban et al., 2018). Binding assays and computational docking studies suggested that CHIL1 binds to demethylxanthohumol and THC to stabilize their ring-opening conformations, circumventing isomerization of THC to naringenin flavanone (Ban et al., 2018). This role of CHIL1 is consistent with the high accumulation of xanthohumol and demethylxanthohumol in hop glandular trichomes, in which almost no THC and naringenin were detected. PT1L is also involved in bitter acid biosynthesis in this plant and physically interacts with another membrane-bound, plastidial, aromatic prenyltransferase, PT2, to form a metabolon that catalyzes the prenylations in the β -bitter acid pathway. In this pathway, PT1L catalyzes the first prenylation and PT2 catalyzes the subsequent two prenylations (Li et al., 2015; Ban et al., 2018). Thus, PT1L might serve as a key scaffold for the biosynthesis of both terpenophenolics (bitter acids and prenylated flavonoids) in hop glandular trichomes.

CONCLUSION AND FUTURE PERSPECTIVES

Interactions between enzymes and proteins in the flavonoid metabolons clearly vary between plant species. This is consistent with the species-dependent structural diversity of flavonoids and points to a role for differential metabolon formation in

producing different structures of flavonoids in a plant species-specific manner. Although specific protein-protein interactions in some flavonoid metabolons have been identified, it is still difficult to grasp the larger picture of flavonoid metabolons and understand how enzymes and proteins dynamically form metabolons to regulate flavonoid biosynthesis. Specifically, the flavonoid metabolon models proposed to date are primarily based on the results of binary interaction analyses and do not describe how three or more enzymes and/or scaffolding proteins are simultaneously and cooperatively associated. Moreover, the protein-protein interactions in flavonoid metabolons remain to be investigated at the atomic level, and their functional significance is yet to be addressed in most important plant taxa. Finally, it is likely that the interactions underlying metabolon formation have evolved to implement higher orders of metabolic functions in the cell, allowing for the structural diversification of flavonoids. Taking full advantage of these important phenomena in synthetic biology has the potential to enhance the efficiency of production of many useful flavonoids in heterologous systems.

AUTHOR CONTRIBUTIONS

TN mainly wrote the manuscript and was responsible for the general opinions stated in the manuscript. All authors reviewed and agreed with the final version of the submitted manuscript.

FUNDING

This study was supported in part by JSPS KAKENHI grant number 18H03938.

REFERENCES

- Achnine, L., Blancaflor, E. B., Rasmussen, S., and Dixon, R. A. (2004). Colocalization of L-phenylalanine ammonia-lyase and cinnamate 4-hydroxylase for metabolic channeling in phenylpropanoid biosynthesis. *Plant Cell* 16, 3098–3109. doi: 10.1105/tpc.104.024406
- An, S., Deng, Y., Tomsho, J. W., Kyoung, M., and Benkovic, S. J. (2010). Microtubule-assisted mechanism for functional metabolic macromolecular complex formation. *Proc. Natl. Acad. Sci. USA* 107, 12872–12876. doi: 10.1073/pnas.1008451107
- Andersen, Ø. M., and Markham, K. R. (2006). *Flavonoids: Chemistry, biochemistry and applications*. (Boca Raton, FL: CRC Press, Taylor & Francis).
- Aoki, T., Akashi, T., and Ayabe, S. (2000). Flavonoids of leguminous plants: structure, biological activity, and biosynthesis. *J. Plant Res.* 113, 475–488. doi: 10.1007/PL00013958
- Arita, M., and Suwa, K. (2008). Search extension transforms wiki into a relational system: a case for flavonoid metabolite database. *BioData Min.* 1:7. doi: 10.1186/1756-0381-1-7
- Austin, M. B., and Noel, J. P. (2003). The chalcone synthase superfamily of type III polyketide synthases. *Nat. Prod. Rep.* 20, 79–110. doi: 10.1039/b100917f
- Ban, Z., Qin, H., Mitchell, A. J., Liu, B., Zhang, F., Weng, J. K., et al. (2018). Noncatalytic chalcone isomerase-fold proteins in *Humulus lupulus* are auxiliary components in prenylated flavonoid biosynthesis. *Proc. Natl. Acad. Sci. USA* 115, E5223–E5232. doi: 10.1073/pnas.1802223115
- Barz, W., and Welle, R. (1992). “Biosynthesis and metabolism of isoflavones and pterocarpan phytoalexins in chickpea, soybean and phytopathogenic fungi” in *Phenolic metabolism in plants*. eds. H. Stafford, and R. Ibrahim, vol. 26 (New York: Plenum Press), 139–164.
- Bomati, E. K., Austin, M. B., Bowman, M. E., Dixon, R. A., and Noel, J. P. (2005). Structural elucidation of chalcone reductase and implications for deoxychalcone biosynthesis. *J. Biol. Chem.* 280, 30496–30503. doi: 10.1074/jbc.M502239200
- Burbulis, I. E., and Winkel-Shirley, B. (1999). Interactions among enzymes of the *Arabidopsis* flavonoid biosynthetic pathway. *Proc. Natl. Acad. Sci. USA* 96, 12929–12934. doi: 10.1073/pnas.96.22.12929
- Cheng, H., Yu, O., and Yu, D. Y. (2008). Polymorphisms of IFS1 and IFS2 gene are associated with isoflavone concentrations in soybean seeds. *Plant Sci.* 175, 505–512. doi: 10.1016/j.plantsci.2008.05.020
- Crosby, K. C., Pietraszewska-Bogiel, A., Gadella, T. W. Jr., and Winkel, B. S. (2011). Förster resonance energy transfer demonstrates a flavonoid metabolon in living plant cells that displays competitive interactions between enzymes. *FEBS Lett.* 585, 2193–2198. doi: 10.1016/j.febslet.2011.05.066
- Dastmalchi, M., Bernards, M. A., and Dhaubhadel, S. (2016). Twin anchors of the soybean isoflavonoid metabolon: evidence for tethering of the complex to the endoplasmic reticulum by IFS and C4H. *Plant J.* 85, 689–706. doi: 10.1111/tpj.13137
- Deguchi, A., Ohno, S., Hosokawa, M., Tatsuzawa, F., and Doi, M. (2013). Endogenous post-transcriptional gene silencing of flavone synthase resulting in high accumulation of anthocyanins in black dahlia cultivars. *Planta* 237, 1325–1335. doi: 10.1007/s00425-013-1848-6
- Diharce, J., Golebiowski, J., Fiorucci, S., and Antonczak, S. (2016). Fine-tuning of microsolvation and hydrogen bond interaction regulates substrate channelling

- in the course of flavonoid biosynthesis. *Phys. Chem. Chem. Phys.* 18, 10337–10345. doi: 10.1039/C5CP05059F
- Fujino, N., Tenma, N., Waki, T., Ito, K., Komatsuzaki, Y., Sugiyama, K., et al. (2018). Physical interactions among flavonoid enzymes in snapdragon and torenia reveal the diversity in the flavonoid metabolon organization of different plant species. *Plant J.* 94, 372–392. doi: 10.1111/tjp.13864
- Fulton, A. B. (1982). How crowded is the cytoplasm? *Cell* 30, 345–347. doi: 10.1016/0092-8674(82)90231-8
- Giege, P., Heazlewood, J. L., Roessner-Tunali, U., Millar, A. H., Fernie, A. R., Leaver, C. J., et al. (2003). Enzymes of glycolysis are functionally associated with the mitochondrion in *Arabidopsis* cells. *Plant Cell* 15, 2140–2151. doi: 10.1105/tpc.012500
- Gou, M., Ran, X., Martin, D. W., and Liu, C. J. (2018). The scaffold proteins of lignin biosynthetic cytochrome P450 enzymes. *Nat. Plants* 4, 299–310. doi: 10.1038/s41477-018-0142-9
- Graham, J. W., Williams, T. C., Morgan, M., Fernie, A. R., Ratcliffe, R. G., and Sweetlove, L. J. (2007). Glycolytic enzymes associate dynamically with mitochondria in response to respiratory demand and support substrate channeling. *Plant Cell* 19, 3723–3738. doi: 10.1105/tpc.107.053371
- Hrazdina, G., and Wagner, G. J. (1985a). Compartmentation of plant phenolic compounds: sites of synthesis and accumulation. *Annu. Proc. Phytochem. Soc. Eur.* 25, 120–133.
- Hrazdina, G., and Wagner, G. J. (1985b). Metabolic pathways as enzyme complexes: evidence for the synthesis of phenylpropanoids and flavonoids on membrane associated enzyme complexes. *Arch. Biochem. Biophys.* 237, 88–100. doi: 10.1016/0003-9861(85)90257-7
- Hrazdina, G., Zobel, A. M., and Hoch, H. C. (1987). Biochemical, immunological, and immunocytochemical evidence for the association of chalcone synthase with endoplasmic reticulum membranes. *Proc. Natl. Acad. Sci. USA* 84, 8966–8970. doi: 10.1073/pnas.84.24.8966
- Jiang, W., Yin, Q., Wu, R., Zheng, G., Liu, J., Dixon, R. A., et al. (2015). Role of a chalcone isomerase-like protein in flavonoid biosynthesis in *Arabidopsis thaliana*. *J. Exp. Bot.* 66, 7165–7179. doi: 10.1093/jxb/erv413
- Jørgensen, K., Rasmussen, A. V., Morant, M., Nielsen, A. H., Bjørnholm, N., Zagrobelny, M., et al. (2005). Metabolon formation and metabolic channeling in the biosynthesis of plant natural products. *Curr. Opin. Plant Biol.* 8, 280–291. doi: 10.1016/j.pbi.2005.03.014
- Kaintz, C., Molitor, C., Thill, J., Kampatsikas, I., Michael, C., Halbwirth, H., et al. (2014). Cloning and functional expression in *E. coli* of a polyphenol oxidase transcript from *Coreopsis grandiflora* involved in aurone formation. *FEBS Lett.* 588, 3417–3426. doi: 10.1016/j.febslet.2014.07.034
- Kaltenbach, M., Burke, J. R., Dindo, M., Pabis, A., Munsberg, F. S., Rabin, A., et al. (2018). Evolution of chalcone isomerase from a noncatalytic ancestor. *Nat. Chem. Biol.* 14, 548–555. doi: 10.1038/s41589-018-0042-3
- Laursen, T., Borch, J., Knudsen, C., Bavishi, K., Torta, F., Martens, H. J., et al. (2016). Characterization of a dynamic metabolon producing the defense compound dhurrin in sorghum. *Science* 354, 890–893. doi: 10.1126/science.aag2347
- Laursen, T., Möller, B. L., and Bassard, J. E. (2015). Plasticity of specialized metabolism as mediated by dynamic metabolons. *Trends Plant Sci.* 20, 20–32. doi: 10.1016/j.tplants.2014.11.002
- Li, H., Ban, Z., Qin, H., Ma, L., King, A. J., and Wang, G. (2015). A heteromeric membrane-bound prenyltransferase complex from hop catalyzes three sequential aromatic prenylations in the bitter acid pathway. *Plant Physiol.* 167, 650–659. doi: 10.1104/pp.114.253682
- Livingstone, J. M., Seguin, P., and Strömvik, M. V. (2010). An *in silico* study of the genes for the isoflavonoid pathway enzymes in soybean reveals novel expressed homologs. *Can. J. Plant Sci.* 90, 453–469. doi: 10.4141/CJPS08214
- Mameda, R., Waki, T., Kawai, Y., Takahashi, S., and Nakayama, T. (2018). Involvement of chalcone reductase in the soybean isoflavone metabolon: identification of GmCHR5, which interacts with 2-hydroxyisoflavanone synthase. *Plant J.* 96, 56–74. doi: 10.1111/tjp.14014
- Moghe, G. D., and Last, R. L. (2015). Something old, something new: conserved enzymes and the evolution of novelty in plant specialized metabolism. *Plant Physiol.* 169, 1512–1523. doi: 10.1104/pp.15.00994
- Morita, Y., Takagi, K., Fukuchi-Mizutani, M., Ishiguro, K., Tanaka, Y., Nitasaka, E., et al. (2014). A chalcone isomerase-like protein enhances flavonoid production and flower pigmentation. *Plant J.* 78, 294–304. doi: 10.1111/tjp.12469
- Nakayama, T., Sato, T., Fukui, Y., Yonekura-Sakakibara, K., Hayashi, H., Tanaka, Y., et al. (2001). Specificity analysis and mechanism of aurone synthesis catalyzed by aureusidin synthase, a polyphenol oxidase homolog responsible for flower coloration. *FEBS Lett.* 499, 107–111. doi: 10.1016/S0014-5793(01)02529-7
- Nakayama, T., Yonekura-Sakakibara, K., Sato, T., Kikuchi, S., Fukui, Y., Fukuchi-Mizutani, M., et al. (2000). Aureusidin synthase: a polyphenol oxidase homolog responsible for flower coloration. *Science* 290, 1163–1166. doi: 10.1126/science.290.5494.1163
- Ngaki, M. N., Louie, G. V., Philippe, R. N., Manning, G., Pojer, F., Bowman, M. E., et al. (2012). Evolution of the chalcone-isomerase fold from fatty-acid binding to stereospecific catalysis. *Nature* 485, 530–533. doi: 10.1038/nature11009
- Oguro, S., Akashi, T., Ayabe, S.-I., Noguchi, H., and Abe, I. (2004). Probing biosynthesis of plant polyketides with synthetic N-acetylcysteamine thioesters. *Biochem. Biophys. Res. Commun.* 325, 561–567. doi: 10.1016/j.bbrc.2004.10.057
- Ono, E., Homma, Y., Horikawa, M., Kunikane-Doi, S., Imai, H., Takahashi, S., et al. (2010). Functional differentiation of the glycosyltransferases that contribute to the chemical diversity of bioactive flavonol glycosides in grapevines (*Vitis vinifera*). *Plant Cell* 22, 2856–2871. doi: 10.1105/tpc.110.074625
- Ono, E., and Nakayama, T. (2007). Molecular breeding of novel yellow flowers by engineering the aurone biosynthetic pathway. *Transgenic Plant J.* 1, 66–80.
- Ovadi, J. (1991). Physiological significance of metabolic channelling. *J. Theor. Biol.* 152, 1–22. doi: 10.1016/S0022-5193(05)80500-4
- Ovadi, J., and Saks, V. (2004). On the origin of intracellular compartmentation and organized metabolic systems. *Mol. Cell. Biochem.* 256, 5–12. doi: 10.1023/B:MCBL.0000009855.14648.2c
- Ovadi, J., and Srere, P. A. (2000). Macromolecular compartmentation and channeling. *Int. Rev. Cytol.* 192, 255–280. doi: 10.1016/S0074-7696(08)60529-X
- Owens, D. K., Alerding, A. B., Crosby, K. C., Bandara, A. B., Westwood, J. H., and Winkel, B. S. (2008). Functional analysis of a predicted flavonol synthase gene family in *Arabidopsis*. *Plant Physiol.* 147, 1046–1061. doi: 10.1104/pp.108.117457
- Ralston, L., Subramanian, S., Matsuno, M., and Yu, O. (2005). Partial reconstruction of flavonoid and isoflavonoid biosynthesis in yeast using soybean type I and type II chalcone isomerases. *Plant Physiol.* 137, 1375–1388. doi: 10.1104/pp.104.054502
- Ralston, L., and Yu, O. (2006). Metabolons involving plant cytochrome P450s. *Phytochem. Rev.* 5, 459–472. doi: 10.1007/s11101-006-9014-4
- Saito, K., Yonekura-Sakakibara, K., Nakabayashi, R., Higashi, Y., Yamazaki, M., Tohge, T., et al. (2013). The flavonoid biosynthetic pathway in *Arabidopsis*: structural and genetic diversity. *Plant Physiol. Biochem.* 72, 21–34. doi: 10.1016/j.plaphy.2013.02.001
- Saks, V., Beraud, N., and Wallimann, T. (2008). Metabolic compartmentation - a system level property of muscle cells: real problems of diffusion in living cells. *Int. J. Mol. Sci.* 9, 751–767. doi: 10.3390/ijms9050751
- Sasaki, N., and Nakayama, T. (2015). Achievements and perspectives in biochemistry concerning anthocyanin modification for blue flower coloration. *Plant Cell Physiol.* 56, 28–40. doi: 10.1093/pcp/pcu097
- Saslowsky, D. E., Warek, U., and Winkel, B. S. (2005). Nuclear localization of flavonoid enzymes in *Arabidopsis*. *J. Biol. Chem.* 280, 23735–23740. doi: 10.1074/jbc.M413506200
- Saslowsky, D., and Winkel-Shirley, B. (2001). Localization of flavonoid enzymes in *Arabidopsis* roots. *Plant J.* 27, 37–48. doi: 10.1046/j.1365-313x.2001.01073.x
- Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., et al. (2010). Genome sequence of the paleopolyploid soybean. *Nature* 463, 178–183. doi: 10.1038/nature08670
- Sepiol, C. J., Yu, J., and Dhaubhadel, S. (2017). Genome-wide identification of chalcone reductase gene family in soybean: insight into root-specific GmCHRs and *Phytophthora sojae* resistance. *Front. Plant Sci.* 8:2073. doi: 10.3389/fpls.2017.02073
- Shih, C. H., Chu, H., Tang, L. K., Sakamoto, W., Maekawa, M., Chu, I. K., et al. (2008). Functional characterization of key structural genes in rice flavonoid biosynthesis. *Planta* 228, 1043–1054. doi: 10.1007/s00425-008-0806-1
- Shimada, N., Aoki, T., Sato, S., Nakamura, Y., Tabata, S., and Ayabe, S. (2003). A cluster of genes encodes the two types of chalcone isomerase involved in the biosynthesis of general flavonoids and legume-specific 5-deoxy(iso) flavonoids in *Lotus japonicus*. *Plant Physiol.* 131, 941–951. doi: 10.1104/pp.004820

- Shimizu, T., Akada, S., Senda, M., Ishikawa, R., Harada, T., Niizeki, M., et al. (1999). Enhanced expression and differential inducibility of soybean chalcone synthase genes by supplemental UV-B in dark-grown seedlings. *Plant Mol. Biol.* 39, 785–795. doi: 10.1023/A:1006124219945
- Shimomura, M., Kanamori, H., Komatsu, S., Namiki, N., Mukai, Y., Kurita, K., et al. (2015). The *Glycine max* cv. Enrei genome for improvement of Japanese soybean cultivars. *Int. J. Genom.* 2015:358127. doi: 10.1155/2015/358127
- Srere, P. A. (2000). Macromolecular interactions: tracing the roots. *Trends Biochem. Sci.* 25, 150–153. doi: 10.1016/S0968-0004(00)01550-4
- Stafford, H. A. (1974). The metabolism of aromatic compounds. *Annu. Rev. Plant Physiol.* 25, 459–486. doi: 10.1146/annurev.pp.25.060174.002331
- Stafford, H. A. (1990). *Flavonoid metabolism*. (Boca Raton, FL: CRC Press).
- Stevens, J. F., and Page, J. E. (2004). Xanthohumol and related prenylflavonoids from hops and beer: to your good health! *Phytochemistry* 65, 1317–1330. doi: 10.1016/j.phytochem.2004.04.025
- Subramanian, S., Stacey, G., and Yu, O. (2006). Endogenous isoflavones are essential for the establishment of symbiosis between soybean and *Bradyrhizobium japonicum*. *Plant J.* 48, 261–273. doi: 10.1111/j.1365-313X.2006.02874.x
- Subramanian, S., Stacey, G., and Yu, O. (2007). Distinct, crucial roles of flavonoids during legume nodulation. *Trends Plant Sci.* 12, 282–285. doi: 10.1016/j.tplants.2007.06.006
- Suss, K. H., Arkona, C., Manteuffel, R., and Adler, K. (1993). Calvin cycle multienzyme complexes are bound to chloroplast thylakoid membranes of higher plants *in situ*. *Proc. Natl. Acad. Sci. USA* 90, 5514–5518. doi: 10.1073/pnas.90.12.5514
- Sweetlove, L. J., and Fernie, A. R. (2013). The spatial organization of metabolism within the plant cell. *Annu. Rev. Plant Biol.* 64, 723–746. doi: 10.1146/annurev-arplant-050312-120233
- Thill, J., Miosic, S., Ahmed, R., Schlangen, K., Muster, G., Stich, K., et al. (2012). ‘Le Rouge et le Noir’: a decline in flavone formation correlates with the rare color of black dahlia (*Dahlia variabilis* hort.) flowers. *BMC Plant Biol.* 12:225. doi: 10.1186/1471-2229-12-225
- Toki, K. (1988). Studies on the flower pigments in snapdragon, *Antirrhinum majus* L. *Bull. Fac. Hortic. MinamiKyushu Univ.* 18, 1–68.
- Tuteja, J. H., Clough, S. J., Chan, W. C., and Vodkin, L. O. (2004). Tissue-specific gene silencing mediated by a naturally occurring chalcone synthase gene cluster in *Glycine max*. *Plant Cell* 16, 819–835. doi: 10.1105/tpc.021352
- Ueyama, Y., Suzuki, K., Fukuchi-Mizutani, M., Fukui, Y., Miyazaki, K., Ohkawa, H., et al. (2002). Molecular and biochemical characterization of torenia flavonoid 3'-hydroxylase and flavone synthase II and modification of flower color by modulating the expression of these genes. *Plant Sci.* 163, 253–263. doi: 10.1016/S0168-9452(02)00098-5
- Verkman, A. S. (2002). Solute and macromolecule diffusion in cellular aqueous compartments. *Trends Biochem. Sci.* 27, 27–33. doi: 10.1016/S0968-0004(01)02003-5
- Waki, T., Yoo, D., Fujino, N., Mameda, R., Denessiouk, K., Yamashita, S., et al. (2016). Identification of protein-protein interactions of isoflavonoid biosynthetic enzymes with 2-hydroxyisoflavanone synthase in soybean (*Glycine max* (L.) Merr.). *Biochem. Biophys. Res. Commun.* 469, 546–551. doi: 10.1016/j.bbrc.2015.12.038
- Watkinson, J. I., Bowerman, P. A., Crosby, K. C., Hildreth, S. B., Helm, R. F., and Winkel, B. S. J. (2018). Identification of MOS9 as an interaction partner for chalcone synthase in the nucleus. *PeerJ* 6:e5598. doi: 10.7717/peerj.5598
- Welle, R., and Grisebach, H. (1988). Isolation of a novel NADPH-dependent reductase which coacts with chalcone synthase in the biosynthesis of 6'-deoxychalcone. *FEBS Lett.* 236, 221–225. doi: 10.1016/0014-5793(88)80318-1
- Welle, R., Schröder, G., Schiltz, E., Grisebach, H., and Schröder, J. (1991). Induced plant responses to pathogen attack. Analysis and heterologous expression of the key enzyme in the biosynthesis of phytoalexins in soybean (*Glycine max* L. Merr. cv. Harosoy 63). *Eur. J. Biochem.* 196, 423–430. doi: 10.1111/j.1432-1033.1991.tb15833.x
- Weng, J. K., and Noel, J. P. (2012). The remarkable pliability and promiscuity of specialized metabolism. *Cold Spring Harb. Symp. Quant. Biol.* 77, 309–320. doi: 10.1101/sqb.2012.77.014787
- Winkel, B. S. (2004). Metabolic channeling in plants. *Annu. Rev. Plant Biol.* 55, 85–107. doi: 10.1146/annurev.arplant.55.031903.141714
- Winkel-Shirley, B. (2001). It takes a garden. How work on diverse plant species has contributed to an understanding of flavonoid metabolism. *Plant Physiol.* 127, 1399–1404. doi: 10.1104/pp.010675
- Wiseman, H. (2006). “Isoflavonoids and human health” in *Flavonoids: Chemistry, biochemistry and applications*. eds. Ø. Andersen, and K. Markham (Boca Raton, FL: CRC Press, Taylor & Francis), 371–396.
- Wu, F., and Minter, S. (2015). Krebs cycle metabolon: structural evidence of substrate channeling revealed by cross-linking and mass spectrometry. *Angew. Chem. Int. Ed.* 54, 1851–1854. doi: 10.1002/anie.201409336
- Wu, F., Pelster, L. N., and Minter, S. D. (2015). Krebs cycle metabolon formation: metabolite concentration gradient enhanced compartmentation of sequential enzymes. *Chem. Commun.* 51, 1244–1247. doi: 10.1039/C4CC08702J

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Nakayama, Takahashi and Waki. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Evolutionary Diversification of Primary Metabolism and Its Contribution to Plant Chemical Diversity

Hiroshi A. Maeda*

Department of Botany, University of Wisconsin–Madison, Madison, WI, United States

OPEN ACCESS

Edited by:

Kazuki Saito,
RIKEN Center for Sustainable
Resource Science (CSRS), Japan

Reviewed by:

Anthony Michael,
UT Southwestern Medical Center,
United States
Philipp Zerbe,
University of California, Davis,
United States

*Correspondence:

Hiroshi A. Maeda
maeda2@wisc.edu

Specialty section:

This article was submitted to
Plant Metabolism
and Chemodiversity,
a section of the journal
Frontiers in Plant Science

Received: 01 May 2019

Accepted: 20 June 2019

Published: 10 July 2019

Citation:

Maeda HA (2019) Evolutionary
Diversification of Primary Metabolism
and Its Contribution to Plant Chemical
Diversity. *Front. Plant Sci.* 10:881.
doi: 10.3389/fpls.2019.00881

Plants produce a diverse array of lineage-specific specialized (secondary) metabolites, which are synthesized from primary metabolites. Plant specialized metabolites play crucial roles in plant adaptation as well as in human nutrition and medicine. Unlike well-documented diversification of plant specialized metabolic enzymes, primary metabolism that provides essential compounds for cellular homeostasis is under strong selection pressure and generally assumed to be conserved across the plant kingdom. Yet, some alterations in primary metabolic pathways have been reported in plants. The biosynthetic pathways of certain amino acids and lipids have been altered in specific plant lineages. Also, two alternative pathways exist in plants for synthesizing primary precursors of the two major classes of plant specialized metabolites, terpenoids and phenylpropanoids. Such primary metabolic diversities likely underlie major evolutionary changes in plant metabolism and chemical diversity by acting as enabling or associated traits for the evolution of specialized metabolic pathways.

Keywords: plant chemical diversity, metabolic enzymes, primary metabolism, specialized metabolism, evolution of plant metabolism, amino acid biosynthesis

INTRODUCTION

Plants produce a diverse array of secondary or specialized metabolites, which play critical roles in plant adaptation under various environmental conditions. These phytochemicals are also widely used in human nutrition and medicine. Nearly one million metabolites are estimated to be produced throughout the plant kingdom (Afendi et al., 2012), though many of them are yet to be discovered. All of these specialized metabolites are synthesized from a certain primary metabolite precursor(s), such as sugars, amino acids, nucleotides, organic acids, and fatty acids, which are essential for maintaining cellular homeostasis and the life of whole organisms. Besides their vital nature, primary metabolic pathways are highly regulated and integrated to complex metabolic networks (Baghalian et al., 2014; Sulpice and McKeown, 2015; Beckers et al., 2016; Filho et al., 2018). Consequently, genes encoding primary metabolic enzymes are subjected to purifying selection and generally considered to be conserved among the plant kingdom, unlike highly diversified

specialized metabolism (Pichersky and Lewinsohn, 2011; Weng et al., 2012; Moghe and Last, 2015; Moore et al., 2019). Yet, some primary metabolic pathways were altered during plant evolution, which had profound impacts on overall plant physiology, metabolism, and adaptation. This review describes examples of primary metabolic diversification in different plant lineages and discusses their potential roles in the evolution of downstream specialized metabolic pathways and plant chemical diversity as enabling or associated traits.

ENABLERS OF EVOLUTIONARY DIVERSIFICATION OF THE PHOTOSYNTHETIC CARBON FIXATION PATHWAYS

One of the most fundamental metabolic pathways of plants, photosynthetic carbon fixation, has been modified in a number of plant lineages to what is known as C_4 photosynthesis and Crassulacean acid metabolism, though the former will be mainly discussed here. Unlike 3-phosphoglycerate (3PGA), a three carbon molecule produced by ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) in C_3 photosynthesis, C_4 photosynthesis initially generates a four carbon molecule, i.e., oxaloacetate, by phosphoenolpyruvate (PEP) carboxylase (PEPC). Oxaloacetate is further converted to malate or aspartate and shuttled from mesophyll to bundle sheath cells, where CO_2 is released for refixation by Rubisco (**Figure 1**) (Langdale, 2011; Sage et al., 2012; Furbank, 2016). This highly intricate mechanism is seemingly maladaptive due to high metabolic costs (e.g., fixing carbon twice, regeneration of PEP), but provides adaptive advantage under arid, warm, and high light conditions by concentrating CO_2 and attenuating the oxygenation side reaction of Rubisco and hence photorespiration (Christin and Osborne, 2014; Sage and Stata, 2015). Thus, besides the decline in atmospheric CO_2 around 30 million years ago (Pagani et al., 2005), such extreme environmental conditions, in which some plants existed, likely acted as an “environmental enabler” for the evolutionary diversification of the photosynthetic carbon fixation, the entry step of plant metabolic pathways.

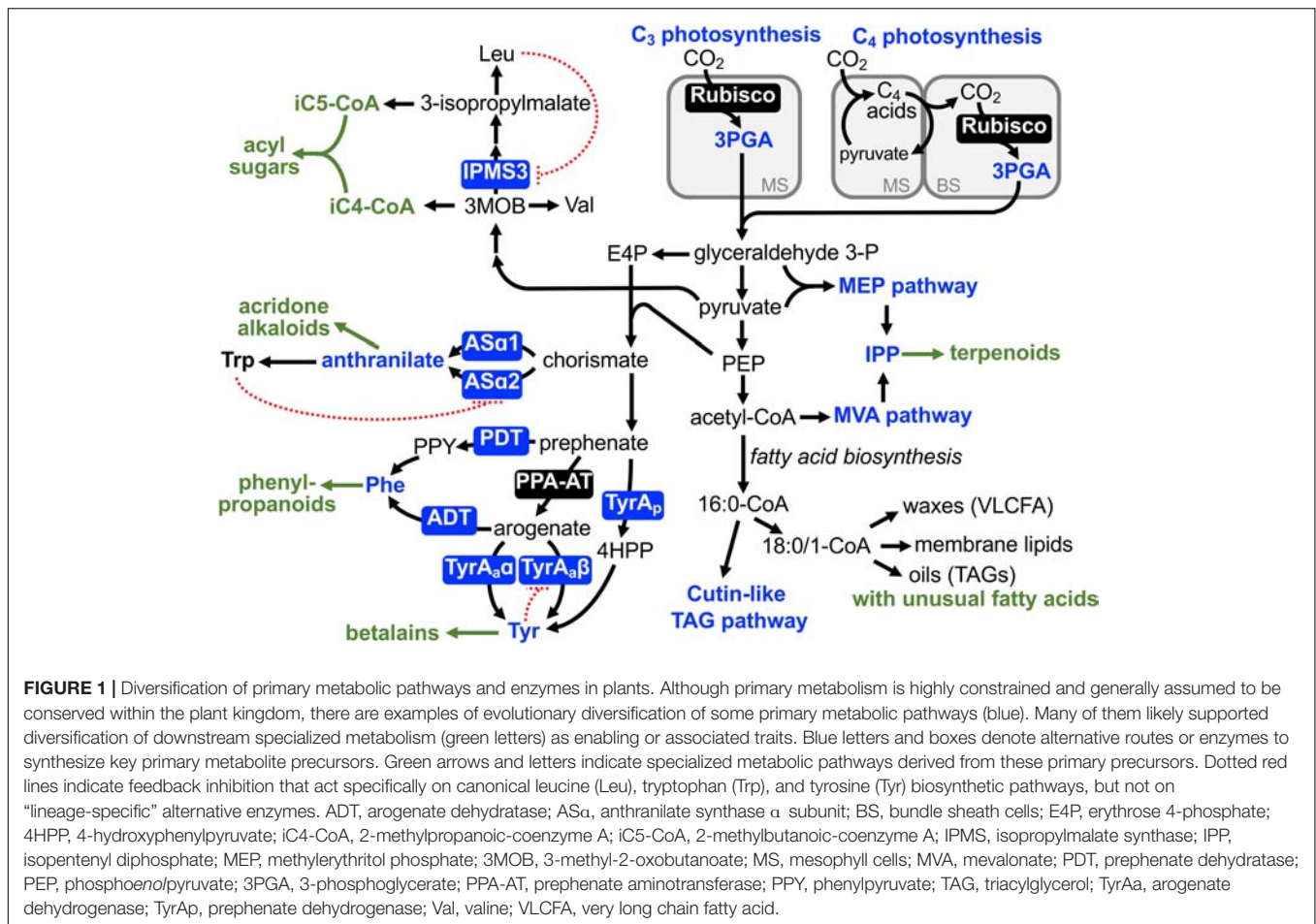
The C_4 photosynthetic pathway evolved more than 60 times independently across the plant phylogeny (Sage et al., 2011, 2012). Notably, C_4 photosynthesis is unevenly distributed across the phylogeny and particularly prevalent in specific plant lineages, such as Poaceae and Caryophyllales (Christin et al., 2009, 2015; Sage et al., 2011). Recent comparative analyses of C_3 and C_4 plants as well as C_3 - C_4 transitory species revealed that the repeated evolution of C_4 photosynthesis was likely facilitated by certain “pre-conditions” or “enabling traits” that emerged or were present in certain plant lineages (Ludwig, 2013; Sage et al., 2014; Heckmann, 2016; Miyake, 2016; Schlüter and Weber, 2016). These enabling traits include “genetic enablers,” such as C_4 -like cell-type specific expression of C_4 enzymes (e.g., PEPC, Williams et al., 2012; Christin et al., 2013a, 2015) and “anatomical enablers,” such as proto-Kranz anatomy (Christin et al., 2013b; Lundgren et al., 2014;

Sage et al., 2014), in C_3 ancestors. These pre-conditions further facilitated emergence of “metabolic enablers,” such as shuttling of photorespiratory glycine from mesophyll to bundle sheath cells acting as CO_2 pump (Sage et al., 2013; Schulze et al., 2013). This so-called C_2 photosynthesis is present in many sister species to C_4 lineages (Sage et al., 2011, 2012; Khoshravesh et al., 2016) and appears to be accompanied by shuttling of other metabolites, such as alanine/pyruvate or aspartate/malate, for balancing of nitrogen between the mesophyll and bundle sheath cells (Mallmann et al., 2014; Schlüter and Weber, 2016). Once these pre-conditions were established, C_4 photosynthesis could evolve relatively easily and thus repeatedly, such as through optimization of kinetic properties of C_4 enzymes (e.g., PEPC) and bundle sheath specific expression of Rubisco (Langdale, 2011; Sage et al., 2012; Furbank, 2016; Reeves et al., 2017). Thus, the combination of environmental, genetic, anatomical, and metabolic enablers allowed astounding alterations in the core primary metabolic pathway, photosynthetic carbon fixation, in certain plant lineages.

DIVERSIFICATION OF AMINO ACID BIOSYNTHETIC PATHWAYS AT THE INTERFACE OF PRIMARY AND SPECIALIZED METABOLIC PATHWAYS

Amino acid biosynthetic pathways not only provide essential protein building blocks but connect central carbon metabolism to a variety of specialized metabolism. Some of these amino acid pathways have diversified in certain plant lineages and likely contributed to the chemical diversity of their downstream specialized metabolism.

Isopropylmalate synthase (IPMS) catalyzes the committed step of leucine biosynthesis (de Kraker et al., 2007). IPMS competes for the 3-methyl-2-oxobutanoate (3MOB) substrate with valine biosynthesis (**Figure 1**) and is typically feedback inhibited by the end product, leucine, through its C-terminal allosteric regulatory domain (Koon et al., 2004; de Kraker and Gershenzon, 2011). Glandular trichomes of Solanaceae plants accumulate insecticidal specialized metabolites, acylsugars, which have various aliphatic acids attached to a sugar backbone (e.g., sucrose, Fan et al., 2019). A wild tomato *Solanum pennellii* and the cultivated tomato, *Solanum lycopersicum*, have 2-methylpropanoic and 3-methylbutanoic acid (iC4 and iC5) acyl chains, which are derived from 3MOB and 3-isopropylmalate, intermediates of valine and leucine metabolism, respectively (**Figure 1**). Analysis of introgression lines between *S. lycopersicum* and *S. pennellii*, followed by expression and biochemical analyses, revealed that the C-terminal regulatory domain of the IPMS3 isoform is truncated in *S. lycopersicum*, making this isoform insensitive to leucine-mediated feedback inhibition (Schillmiller et al., 2010; Ning et al., 2015). In contrast, the IPMS3 isoform of *S. pennellii* is further truncated into its catalytic domain and has lost the enzyme activity. Thus, the de-regulated and inactive IPMS3 in *S. lycopersicum* and *S. pennellii* directs more carbon flow



toward leucine and valine metabolism, respectively. Having the broad substrate specificity of downstream acyl-CoA-dependent acyltransferase (Schillmiller et al., 2015), increased availability of 3MOB and 3-isopropylmalate contributes to the formation of iC4 and iC5 acylsugars, respectively. Brassicaceae species including *Arabidopsis thaliana* also has a truncated IPMS homolog but with point mutations that alter substrate specificity to now function as methylthioalkylmalate synthase in the initial step of methionine-derived glucosinolate biosynthesis (de Kraker and Gershenzon, 2011). Unlike the latter example of recruitment of specialized metabolic enzymes from primary metabolism, as discussed in previous reviews (Weng, 2014; Moghe and Last, 2015), the study by Ning et al. (2015) revealed a role of altered branch chain amino acid biosynthesis in the acyl chain diversity of acylsugars in the *Solanum* genus.

Anthranilate synthase (AS) catalyzes the committed step of biosynthesis of an aromatic amino acid, L-tryptophan, and its enzyme activity is strictly regulated through feedback inhibition of one of the AS enzyme complex, ASα, by tryptophan (Romero et al., 1995; Li and Last, 1996). Two copies of ASα genes, ASα1 and ASα2, were found in *Ruta graveolens* (the Rutaceae family) that uses anthranilate to produce unique specialized metabolites, acridone alkaloids (Bohlmann et al., 1995). While

ASα2 was constitutively expressed, ASα1 was induced under elicitor treatment, which stimulates the accumulation of acridone alkaloids. Interestingly, the ASα1 enzyme was much more resistant than ASα2 to the tryptophan-mediated feedback inhibition, suggesting that the expression of the de-regulated ASα1 enzyme allowed elevated accumulation of the anthranilate precursor and hence efficient production of the downstream specialized metabolites, acridone alkaloids, in this unique plant lineage (Bohlmann et al., 1996; **Figure 1**). A naturally occurring feedback-insensitive ASα enzyme has also been identified in *Nicotiana tabacum* (the Solanaceae family, Song et al., 1998), but its *in planta* function is currently unknown. Further evolutionary analyses across the Rutaceae family can evaluate if the increased availability of anthranilate served as an enabling trait for later evolution of acridone alkaloid biosynthesis. Alternatively, the de-regulated ASα1 might have evolved after the emergence of the acridone alkaloid pathway as an associated trait and further elevated the alkaloid production.

L-Tyrosine is another aromatic amino acid required for protein synthesis but also used to produce diverse plant natural products, such as tocochromanols, benzylisoquinoline alkaloids, cyanogenic glycosides (e.g., dhurrin), and rosmarinic acids (Schenck and Maeda, 2018). Tyrosine is typically produced

via arogenate dehydrogenase (TyrA_a) that is localized within the plastids (Rippert et al., 2009; Wang et al., 2016) and strongly feedback inhibited by tyrosine (**Figure 1**; Connelly and Conn, 1986; Rippert and Matringe, 2002a,b). Recent studies, however, uncovered diversification of the tyrosine biosynthetic pathways in different plant lineages. In addition to the highly regulated plastidic TyrA_a-mediated pathway, many legumes including *Glycine max* (soybean) and *Medicago truncatula* have an additional tyrosine biosynthetic pathway mediated by prephenate dehydrogenase (TyrA_p) (Rubin and Jensen, 1979; Schenck et al., 2015), which is often found in microbes (Bonner and Jensen, 1987; Bonner et al., 2008; Schenck et al., 2017b). Notably, these legume TyrA_p enzymes are localized outside of the plastids and completely insensitive to feedback inhibition by tyrosine (Schenck et al., 2015, 2017a), suggesting that the alternative tyrosine pathway is physically separated from the canonical plastidic pathway and escaped feedback inhibition by tyrosine (**Figure 1**). While the metabolic and physiological functions of the alternative cytosolic TyrA_p pathway in legumes is largely unknown, some legumes accumulate very high levels of tyrosine and tyrosine-derived compounds (e.g., L-DOPA in *Mucuna pruriens*, Wichers et al., 1993; Lokvam et al., 2006). A recent study found that the expression of gene encoding the tyrosine-insensitive TyrA_p enzyme is elevated in *Inga* species that accumulate tyrosine and its derived secondary metabolites (e.g., tyrosine-gallates) at 5 to 20% of seedling dry weight (Coley et al., 2019). Thus, the presence of the feedback-insensitive TyrA_p enzyme in the legume family likely provided a unique pre-condition that enabled increased tyrosine biosynthetic activity and hyper-accumulation of tyrosine-derived compounds in this specific genus of legumes.

Betalains are red to yellow alkaloid pigments uniquely produced in the plant order Caryophyllales, which include *Beta vulgaris* (beet), spinach, quinoa, and cactus. Betalain pigments are derived from tyrosine and replaced more ubiquitous red to purple anthocyanin pigments derived from phenylalanine in many Caryophyllales species (Tanaka et al., 2008; Brockington et al., 2011; Polturak and Aharoni, 2018; **Figure 1**). Like Arabidopsis and unlike legumes, Caryophyllales species only have arogenate-specific TyrA_a enzymes; however, one TyrA_a isoform (TyrA_aα) exhibits relaxed sensitivity to tyrosine inhibition (Lopez-Nieves et al., 2018; **Figure 1**). The presence of the de-regulated TyrA_aα enzymes positively and negatively correlates with those of betalain and anthocyanin pigmentation, respectively, across Caryophyllales. Evolutionary analyses, by utilizing transcriptome data of over one hundred Caryophyllales species (Brockington et al., 2015), revealed that the de-regulated TyrA_aα enzymes emerged *before* the evolution of the betalain biosynthetic pathway (Lopez-Nieves et al., 2018). Thus, the enhanced supply of the tyrosine precursor, due to relaxed regulation of the TyrA_a enzyme, likely acted as a metabolic enabler for the subsequent evolution of a novel downstream specialized metabolic pathway, betalain biosynthesis, in this specific plant order (**Figure 1**). Further evolutionary analyses of associated genes and enzymes involved in the betalain pathway and the competing phenylalanine

and phenylpropanoid pathways will provide novel insight into how primary and specialized metabolism evolved coordinately in a macroevolutionary scale beyond the levels of species and genera.

ANCIENT DIVERSIFICATION OF IPP AND PHENYLALANINE BIOSYNTHETIC PATHWAYS IN PLANTAE

In the ancient history of Plantae, alternative primary metabolic pathways evolved and likely contributed to later evolution of plant specialized metabolism and chemical diversity. Terpenoids and phenylpropanoids are the two major classes of plant natural products, which are synthesized from the primary metabolite precursors, isopentenyl pyrophosphate (IPP) and phenylalanine, respectively (McGarvey and Croteau, 1995; Gershenzon and Dudareva, 2007; Vogt, 2010; Tohge et al., 2013). Notably, plants possess two alternative pathways to synthesize IPP and phenylalanine.

In addition to sterols and quinones, plants use IPP to synthesize photosynthetic pigments (chlorophylls, carotenoids), plant hormones (brassinosteroids, abscisic acid, gibberellins), and a diverse array of terpenoid compounds (McGarvey and Croteau, 1995; Gershenzon and Dudareva, 2007; Tholl, 2015). Such a high demand of IPP for synthesis of diverse terpenoid compounds in plants is supported by the two alternative IPP biosynthetic pathways, the methylerythritol phosphate (MEP) and mevalonate (MVA) pathways, which take place in the plastidic and extra-plastidic subcellular compartments, respectively (Vranová et al., 2013; Rodríguez-Concepción and Boronat, 2015). The MEP pathway utilizes glyceraldehyde 3-phosphate derived from the pentose phosphate pathways in the plastids and hence can draw carbon flux directly from photosynthetic carbon fixation (**Figure 1**). While the MVA pathway appears to be an ancestral pathway that evolved in all three domains of life (i.e., eukaryotes, archaea, and most bacteria) or in their last universal ancestor (i.e., cenancestor) (Lombard and Moreira, 2011), the plastidic MEP pathway has mosaic evolutionary origins (Lange et al., 2000; Matsuzaki et al., 2008). A common ancestor of plastid bearing eukaryotes likely acquired MEP pathway enzymes from various bacterial ancestors (i.e., cyanobacteria, α-proteobacteria, Chlamydia) through horizontal gene transfers (Matsuzaki et al., 2008) and the MEP pathway was vertically transmitted to the descendants, the entire Plantae including algae and plants.

L-Phenylalanine is the primary metabolite precursor of phenylpropanoids and is synthesized via two alternative pathways in plants (Tzin and Galili, 2010; Maeda and Dudareva, 2012; Yoo et al., 2013; Qian et al., 2019). In many microbes, phenylalanine is synthesized via the phenylpyruvate intermediate, catalyzed by prephenate dehydratase (PDT) and phenylpyruvate aminotransferase (**Figure 1**) (Bentley, 1990). Although an analogous phenylpyruvate pathway also exists in the plant cytosol (Yoo et al., 2013; Qian et al., 2019), plants synthesize phenylalanine mainly in the plastids via the

L-arogenate intermediate: prephenate is first transaminated by prephenate aminotransferase (PPA-AT) to arogenate (Graindorge et al., 2010; Dal Cin et al., 2011; Maeda et al., 2011), which is then converted to phenylalanine by arogenate dehydratase (ADT; Siehl and Conn, 1988; Cho et al., 2007; Maeda et al., 2010; **Figure 1**). Evolutionary analyses of the PPA-AT and ADT enzymes suggested that an ancestor of green algae and land plants appear to have acquired both of these two enzymes from an ancestor of Chlorobi/Bacteroidetes bacteria, likely through horizontal gene transfer (Dornfeld et al., 2014). Some cyanobacteria also have PPA-AT enzymes but with a distinct evolutionary origin from those of plants and Chlorobi/Bacteroidetes bacteria (Graindorge et al., 2014; Giustini et al., 2019). Thus, these dual primary metabolic pathways of isoprenoid and phenylalanine biosynthesis appear to have evolved in a common ancestor of Plantae. Although evolutionary analyses of such deep phylogenetic nodes are challenging, these dual precursor supply pathways potentially served as metabolic enablers for the evolutionary expansion of terpenoids and phenylpropanoids, the hallmarks of chemical diversity uniquely seen in the plant kingdom today.

DIVERSIFICATION OF LIPID METABOLISM IN PLANTS

Notable chemical diversity also exists in plant lipid metabolism (Badami and Patil, 1980; Ohlrogge et al., 2018), which makes the boundary of primary and specialized (secondary) metabolism difficult to define. Besides major acyl chains (e.g., oleic 18:1, linolenic 18:3) found in most plant lipids, some plants produce unusual fatty acids: For example, oils of castor (*Ricinus communis*, Euphorbiaceae family) and *Vernonia galamensis* (Asteraceae family) consist of primarily (80–90%) hydroxylated and epoxy fatty acids, respectively (Canvin, 1963; Ayorinde et al., 1990). Also, diverse acetylenic natural products having a carbon-carbon triple bond(s) or alkynyl functional group can be produced by modification of the fatty acid precursors (Minto and Blacklock, 2008; Negri, 2015). The production of these hydroxylated fatty acids and polyacetylenes are mediated by divergent fatty acid desaturases with altered product specificities and catalytic properties (van de Loo et al., 1995; Broun et al., 1998; Liu et al., 1998; Broadwater et al., 2002; Minto and Blacklock, 2008; Negri, 2015). Tremendous diversity of cuticular waxes has been also documented across the plant kingdom likely due to the presence of specialized acyl chain elongation and modifying enzymes (Jetter et al., 2007; Busta and Jetter, 2018).

Recent studies also revealed an intriguing alteration in the core lipid metabolic pathway, triacylglycerol (TAG) biosynthesis, in a specific plant lineage. The fruits of Bayberry (*Myrica pensylvanica*, Myricaceae family) accumulate abundant and unusual extracellular glycerolipids: TAG, diacylglycerol (DAG), and monoacylglycerol with completely saturated acyl chains at up to 30% of fruit dry weight (Harlow et al., 1965; Simpson and Ohlrogge, 2016). This unique surface wax attracts birds for seed dispersal and is used for making scented

candles (Fordham, 1983). Fleshy fruits of oil palm, olive, and avocado also accumulate a large quantity of glycerolipids but intracellularly and by upregulating conventional fatty acid and TAG biosynthetic pathways (Bourgis et al., 2011; Kilaru et al., 2015). In contrast, a novel TAG biosynthetic pathway evolved in Bayberry through “re-purposing” genes and enzymes involved in cutin biosynthesis by altering their gene expression (Simpson and Ohlrogge, 2016; Simpson et al., 2016). These alterations include elevated expression of genes encoding the G subfamily of ABC (ABCG) transporters and lipid transporter proteins likely required for lipid transport across cell membranes and walls, respectively, which will allow extracellular formation of TAG (Simpson et al., 2016). It will be interesting to examine how such reprogramming of existing lipid metabolic pathways occur in a step-wise manner during evolution, which will provide useful information for engineering other plants to produce and secrete abundant extracellular glycerolipids.

SUMMARY AND PERSPECTIVE

Although not as frequent as those of specialized metabolism, accumulating evidence indicates that pathways and enzymes of primary metabolism can be diversified during the plant evolution. Such relatively rare alterations in primary metabolism likely contributed to major evolutionary innovations in the plant kingdom, including the evolution of downstream specialized metabolic pathways and hence plant chemical diversity. Some alterations in primary metabolism appear to have acted as enabling traits for the evolution of novel specialized metabolism, at least in the case of de-regulated tyrosine biosynthesis in Caryophyllales that preceded the emergence of betalain pigmentation (Lopez-Nieves et al., 2018). In other instances, primary metabolic alterations likely co-evolved with and support efficient operation of specialized metabolic pathways. It remains to be examined how prevalent the phenomenon is beyond the pathways and plant lineages that have been examined so far and what impacts such primary metabolic diversification had on overall metabolism, physiology, and environmental adaption of diverse plant species. Another intriguing question is how seemingly maladaptive alterations in highly conserved and constrained primary metabolism were maintained in certain plant lineages, especially until the emergence of a new downstream pathway which might have eventually provided adaptive advantage. What are the environmental, anatomical, and genetic enablers underlying primary metabolic diversification? In the case of tomato feedback-insensitive IPMS and legume TyrA_p enzymes, their specific expression in the apical trichome cells (Ning et al., 2015) and extra-plastidic subcellular compartment (Schenck et al., 2015) likely allow minimal disturbance to *de novo* biosynthesis of branch chain and aromatic amino acids, respectively. Further addressing these questions will lead to broader understanding of the evolution of plant metabolism at a macroevolutionary scale. The acquired knowledge of primary metabolic diversification and its underlying genetic and biochemical basis will also allow us to redesign plant metabolism in a holistic manner from primary to specialized metabolism.

AUTHOR CONTRIBUTIONS

HM wrote the manuscript.

FUNDING

This work was supported by the National Science Foundation grant IOS-1354971 and the Agriculture and Food Research

Initiative Competitive grant (2015-67013-22955) from the USDA National Institute of Food and Agriculture to HM.

ACKNOWLEDGMENTS

I would like to thank Dr. Luke Busta for helpful discussion and suggestions.

REFERENCES

- Afendi, F. M., Okada, T., Yamazaki, M., Hirai-Morita, A., Nakamura, Y., Nakamura, K., et al. (2012). KNApSAC family databases: integrated metabolite-plant species databases for multifaceted plant research. *Plant Cell Physiol.* 53:e1. doi: 10.1093/pcp/pcr165
- Ayorinde, F. O., Butler, B. D., and Clayton, M. T. (1990). *Vernonia galamensis*: a rich source of epoxy acid. *J. Am. Oil Chem. Soc.* 67, 844–845. doi: 10.1007/BF02540503
- Badami, R. C., and Patil, K. B. (1980). Structure and occurrence of unusual fatty acids in minor seed oils. *Prog. Lipid Res.* 19, 119–153. doi: 10.1016/0163-7827(80)90002-8
- Baghalian, K., Hajirezaei, M. R., and Schreiber, F. (2014). Plant metabolic modeling: achieving new insight into metabolism and metabolic engineering. *Plant Cell* 26, 3847–3866. doi: 10.1105/tpc.114.130328
- Beckers, V., Dersch, L. M., Lotz, K., Melzer, G., Blasing, O. E., Fuchs, R., et al. (2016). *In silico* metabolic network analysis of Arabidopsis leaves. *BMC Syst. Biol.* 10:102. doi: 10.1186/s12918-016-0347-3
- Bentley, R. (1990). The shikimate pathway - A metabolic tree with many branches. *Crit. Rev. Biochem. Mol. Biol.* 25, 307–384. doi: 10.3109/10409239009090615
- Bohlmann, J., DeLuca, V., Eilert, U., and Martin, W. (1995). Purification and cDNA cloning of anthranilate synthase from *Ruta graveolens*: modes of expression and properties of native and recombinant enzymes. *Plant J.* 7, 491–501. doi: 10.1046/j.1365-3113.1995.7030491.x
- Bohlmann, J., Lins, T., Martin, W., and Eilert, U. (1996). Anthranilate synthase from *Ruta graveolens*. Duplicated AS alpha genes encode tryptophan-sensitive and tryptophan-insensitive isoenzymes specific to amino acid and alkaloid biosynthesis. *Plant Physiol.* 111, 507–514. doi: 10.1104/pp.111.2.507
- Bonner, C., and Jensen, R. (1987). Arogenate dehydrogenase. *Methods Enzymol.* 142, 488–494. doi: 10.1016/s0076-6879(87)42060-0
- Bonner, C. A., Disz, T., Hwang, K., Song, J., Vonstein, V., Overbeek, R., et al. (2008). Cohesion group approach for evolutionary analysis of TyrA, a protein family with wide-ranging substrate specificities. *Microbiol. Mol. Biol. Rev.* 72, 13–53. doi: 10.1128/MMBR.00026-27
- Bourgis, F., Kilaru, A., Cao, X., Ngando-Ebongue, G. F., Drira, N., Ohlrogge, J. B., et al. (2011). Comparative transcriptome and metabolite analysis of oil palm and date palm mesocarp that differ dramatically in carbon partitioning. *Proc. Natl. Acad. Sci. U.S.A.* 108, 12527–12532. doi: 10.1073/pnas.1106502108
- Broadwater, J. A., Whittle, E., and Shanklin, J. (2002). Desaturation and hydroxylation. Residues 148 and 324 of Arabidopsis FAD2, in addition to substrate chain length, exert a major influence in partitioning of catalytic specificity. *J. Biol. Chem.* 277, 15613–15620. doi: 10.1074/jbc.M200231200
- Brockington, S. F., Walker, R. H., Glover, B. J., Soltis, P. S., and Soltis, D. E. (2011). Complex pigment evolution in the Caryophyllales. *New Phytol.* 190, 854–864. doi: 10.1111/j.1469-8137.2011.03687.x
- Brockington, S. F., Yang, Y., Gandia-Herrero, F., Covshoff, S., Hibberd, J. M., Sage, R. F., et al. (2015). Lineage-specific gene radiations underlie the evolution of novel betalain pigmentation in Caryophyllales. *New Phytol.* 207, 1170–1180. doi: 10.1111/nph.13441
- Broun, P., Shanklin, J., Whittle, E., and Somerville, C. (1998). Catalytic plasticity of fatty acid modification enzymes underlying chemical diversity of plant lipids. *Science* 282, 1315–1317. doi: 10.1126/science.282.5392.1315
- Busta, L., and Jetter, R. (2018). Moving beyond the ubiquitous: the diversity and biosynthesis of specialty compounds in plant cuticular waxes. *Phytochem. Rev.* 17, 1275–1304. doi: 10.1007/s11101-017-9542-9540
- Canvin, D. (1963). Formation of oil in seed of *Ricinus communis* L. *Can. J. Biochem. Physiol.* 41, 1879–1885. doi: 10.1139/y63-214
- Cho, M. H., Corea, O. R. A., Yang, H., Bedgar, D. L., Laskar, D. D., Anterola, A. M., et al. (2007). Phenylalanine biosynthesis in *Arabidopsis thaliana* - Identification and characterization of arogenate dehydratases. *J. Biol. Chem.* 282, 30827–30835. doi: 10.1074/jbc.M702662200
- Christin, P. A., Arakaki, M., Osborne, C. P., and Edwards, E. J. (2015). Genetic enablers underlying the clustered evolutionary origins of C4 photosynthesis in angiosperms. *Mol. Biol. Evol.* 32, 846–858. doi: 10.1093/molbev/msu410
- Christin, P.-A., Boxall, S. F., Gregory, R., Edwards, E. J., Hartwell, J., and Osborne, C. P. (2013a). Parallel recruitment of multiple genes into C4 photosynthesis. *Genome Biol. Evol.* 5, 2174–2187. doi: 10.1093/gbe/evt168
- Christin, P.-A., Osborne, C. P., Chatelet, D. S., Columbus, J. T., Besnard, G., Hodkinson, T. R., et al. (2013b). Anatomical enablers and the evolution of C4 photosynthesis in grasses. *Proc. Natl. Acad. Sci. U.S.A.* 110, 1381–1386. doi: 10.1073/pnas.1216777110
- Christin, P.-A., and Osborne, C. P. (2014). The evolutionary ecology of C4 plants. *New Phytol.* 204, 765–781. doi: 10.1111/nph.13033
- Christin, P.-A., Salamin, N., Kellogg, E. A., Vicentini, A., and Besnard, G. (2009). Integrating phylogeny into studies of C4 variation in the grasses. *Plant Physiol.* 149, 82–87. doi: 10.1104/pp.108.128553
- Coley, P. D., Endara, M.-J., Ghabash, G., Kidner, C. A., Nicholls, J. A., Pennington, R. T., et al. (2019). Macroevolutionary patterns in overexpression of tyrosine: an anti-herbivore defence in a speciose tropical tree genus, *Inga* (Fabaceae). *J. Ecol.* 107, 1620–1632. doi: 10.1111/1365-2745.13208
- Connelly, J. A., and Conn, E. E. (1986). Tyrosine biosynthesis in *Sorghum bicolor*: isolation and regulatory properties of arogenate dehydrogenase. *Z. Naturforsch.* C 41, 69–78. doi: 10.1515/znc-1986-1-212
- Dal Cin, V., Tieman, D. M., Tohge, T., McQuinn, R., de Vos, R. C. H., Osorio, S., et al. (2011). Identification of genes in the phenylalanine metabolic pathway by ectopic expression of a MYB transcription factor in tomato fruit. *Plant Cell* 23, 2738–2753. doi: 10.1105/tpc.111.086975
- de Kraker, J.-W., and Gershenzon, J. (2011). From amino acid to glucosinolate biosynthesis: protein sequence changes in the evolution of methylthioalkylmalate synthase in *Arabidopsis*. *Plant Cell* 23, 38–53. doi: 10.1105/tpc.110.079269
- de Kraker, J.-W., Luck, K., Textor, S., Tokuhisa, J. G., and Gershenzon, J. (2007). Two Arabidopsis genes (IPMS1 and IPMS2) encode isopropylmalate synthase, the branchpoint step in the biosynthesis of leucine. *Plant Physiol.* 143, 970–986. doi: 10.1104/pp.106.085555
- Dornfeld, C., Weisberg, A. J., Ritesh, K. C., Dudareva, N., Jelesko, J. G., and Maeda, H. A. (2014). Phylobiochemical characterization of class-Ib aspartate/prephenate aminotransferases reveals evolution of the plant arogenate phenylalanine pathway. *Plant Cell* 26, 3101–3114. doi: 10.1105/tpc.114.127407
- Fan, P., Leong, B. J., and Last, R. L. (2019). Tip of the trichome: evolution of acylsugar metabolic diversity in Solanaceae. *Curr. Opin. Plant Biol.* 49, 8–16. doi: 10.1016/j.pbi.2019.03.005
- Filho, H. A., Machicao, J., and Bruno, O. M. (2018). A hierarchical model of metabolic machinery based on the kcore decomposition of plant metabolic networks. *PLoS One* 13:e0195843. doi: 10.1371/journal.pone.0195843
- Fordham, A. J. (1983). Of Birds and Bayberries: seed dispersal and propagation of three *Myrica* species. *Arnoldia* 43, 20–23.
- Furbank, R. T. (2016). Walking the C4 pathway: past, present, and future. *J. Exp. Bot.* 67, 4057–4066. doi: 10.1093/jxb/erw161

- Gershenzon, J., and Dudareva, N. (2007). The function of terpene natural products in the natural world. *Nat. Chem. Biol.* 3, 408–414. doi: 10.1038/nchembio.2007.5
- Giustini, C., Graindorge, M., Cobessi, D., Crouzy, S., Robin, A., Curien, G., et al. (2019). Tyrosine metabolism: identification of a key residue in the acquisition of prephenate aminotransferase activity by β aspartate aminotransferase. *FEBS J.* 286, 2118–2134. doi: 10.1111/febs.14789
- Graindorge, M., Giustini, C., Jacomin, A. C., Kraut, A., Curien, G., and Matringe, M. (2010). Identification of a plant gene encoding glutamate/aspartate-prephenate aminotransferase: the last homeless enzyme of aromatic amino acids biosynthesis. *FEBS Lett.* 584, 4357–4360. doi: 10.1016/j.febslet.2010.09.037
- Graindorge, M., Giustini, C., Kraut, A., Moyet, L., Curien, G., and Matringe, M. (2014). Three different classes of aminotransferases evolved prephenate aminotransferase functionality in aroenate-competent microorganisms. *J. Biol. Chem.* 289, 3198–3208. doi: 10.1074/jbc.M113.486480
- Harlow, R. D., Litchfield, C., Fu, H.-C., and Reiser, R. (1965). The triglyceride composition of *Myrica carolinensis* fruit coat fat (bayberry tallow). *J. Am. Oil Chem. Soc.* 42, 747–750. doi: 10.1007/BF02631853
- Heckmann, D. (2016). C4 photosynthesis evolution: the conditional Mt. Fuji. *Curr. Opin. Plant Biol.* 31, 149–154. doi: 10.1016/j.pbi.2016.04.008
- Jetter, R., Kunst, L., and Samuels, A. L. (2007). “Composition of Plant Cuticular Waxes,” in *Annual Plant Reviews Volume 23: Biology of the Plant Cuticle*, eds M. Riederer and C. Muller (Hoboken, NJ: John Wiley & Sons, Ltd.), 145–181. doi: 10.1002/9781119312994.apr0232
- Khoshravesh, R., Stinson, C. R., Stata, M., Busch, F. A., Sage, R. F., Ludwig, M., et al. (2016). C3–C4 intermediacy in grasses: organelle enrichment and distribution, glycine decarboxylase expression, and the rise of C2 photosynthesis. *J. Exp. Bot.* 67, 3065–3078. doi: 10.1093/jxb/erw150
- Kilaru, A., Cao, X., Dabbs, P. B., Sung, H.-J., Rahman, M. M., Thrower, N., et al. (2015). Oil biosynthesis in a basal angiosperm: transcriptome analysis of *Persea americana* mesocarp. *BMC Plant Biol.* 15:203. doi: 10.1186/s12870-015-0586-582
- Koon, N., Squire, C. J., and Baker, E. N. (2004). Crystal structure of LeuA from *Mycobacterium tuberculosis*, a key enzyme in leucine biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* 101, 8295–8300. doi: 10.1073/pnas.0400820101
- Langdale, J. A. (2011). C4 cycles: past, present, and future research on C4 Photosynthesis. *Plant Cell* 23, 3879–3892. doi: 10.1105/tpc.111.092098
- Lange, B. M., Rujan, T., Martin, W., and Croteau, R. (2000). Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. *Proc. Natl. Acad. Sci. U.S.A.* 97, 13172–13177. doi: 10.1073/pnas.240454797
- Li, J., and Last, R. L. (1996). The *Arabidopsis thaliana* *trp5* mutant has a feedback-resistant anthranilate synthase and elevated soluble tryptophan. *Plant Physiol.* 110, 51–59. doi: 10.1104/pp.110.1.51
- Liu, L., Hammond, E. G., and Nikolau, B. J. (1998). In vivo studies of the biosynthesis of vernolic acid in the seed of *Vernonia galamensis*. *Lipids* 33, 1217–1221. doi: 10.1007/s11745-998-0326-3
- Lokvam, J., Brenes-Arguedas, T., Lee, J. S., Coley, P. D., and Kursar, T. A. (2006). Allelochemic function for a primary metabolite: the case of L-tyrosine hyperproduction in *Inga umbellifera* (Fabaceae). *Am. J. Bot.* 93, 1109–1115. doi: 10.3732/ajb.93.8.1109
- Lombard, J., and Moreira, D. (2011). Origins and early evolution of the mevalonate pathway of isoprenoid biosynthesis in the three domains of life. *Mol. Biol. Evol.* 28, 87–99. doi: 10.1093/molbev/msq177
- Lopez-Nieves, S., Yang, Y., Timoneda, A., Wang, M., Feng, T., Smith, S. A., et al. (2018). Relaxation of tyrosine pathway regulation underlies the evolution of betalain pigmentation in Caryophyllales. *New Phytol.* 217, 896–908. doi: 10.1111/nph.14822
- Ludwig, M. (2013). Evolution of the C4 photosynthetic pathway: events at the cellular and molecular levels. *Photosyn. Res.* 117, 147–161. doi: 10.1007/s11220-013-9853-y
- Lundgren, M. R., Osborne, C. P., and Christin, P. A. (2014). Deconstructing Kranz anatomy to understand C4 evolution. *J. Exp. Bot.* 65, 3357–3369. doi: 10.1093/jxb/eru186
- Maeda, H., and Dudareva, N. (2012). The shikimate pathway and aromatic amino acid biosynthesis in plants. *Annu. Rev. Plant Biol.* 63, 73–105. doi: 10.1146/annurev-arplant-042811-105439
- Maeda, H., Shasany, A. K., Schnepf, J., Orlova, I., Taguchi, G., Cooper, B. R., et al. (2010). RNAi suppression of *Arogenate Dehydratase1* reveals that phenylalanine is synthesized predominantly via the aroenate pathway in petunia petals. *Plant Cell* 22, 832–849. doi: 10.1105/tpc.109.073247
- Maeda, H., Yoo, H., and Dudareva, N. (2011). Prephenate aminotransferase directs plant phenylalanine biosynthesis via aroenate. *Nat. Chem. Biol.* 7, 19–21. doi: 10.1038/nchembio.485
- Mallmann, J., Heckmann, D., Bräutigam, A., Lercher, M. J., Weber, A. P. M., Westhoff, P., et al. (2014). The role of photorespiration during the evolution of C4 photosynthesis in the genus *Flaveria*. *eLife* 3, e02478. doi: 10.7554/eLife.02478
- Matsuzaki, M., Kuroiwa, H., Kuroiwa, T., Kita, K., and Nozaki, H. (2008). A cryptic algal group unveiled: a plastid biosynthesis pathway in the oyster parasite *Perkinsus marinus*. *Mol. Biol. Evol.* 25, 1167–1179. doi: 10.1093/molbev/msn064
- McGarvey, D. J., and Croteau, R. (1995). Terpenoid metabolism. *Plant Cell* 7, 1015–1026. doi: 10.1105/tpc.7.7.1015
- Minto, R. E., and Blacklock, B. J. (2008). Biosynthesis and function of polyacetylenes and allied natural products. *Prog. Lipid Res.* 47, 233–306. doi: 10.1016/j.plipres.2008.02.002
- Miyake, H. (2016). Starch accumulation in the bundle sheaths of C3 plants: a possible pre-condition for C4 Photosynthesis. *Plant Cell Physiol.* 57, 890–896. doi: 10.1093/pcp/pcw046
- Moghe, G. D., and Last, R. L. (2015). Something old, something new: conserved enzymes and the evolution of novelty in plant specialized metabolism. *Plant Physiol.* 169, 1512–1523. doi: 10.1104/pp.15.00994
- Moore, B. M., Wang, P., Fan, P., Leong, B., Schenck, C. A., Lloyd, J. P., et al. (2019). Robust predictions of specialized metabolism genes through machine learning. *Proc. Natl. Acad. Sci. U.S.A.* 116, 2344–2353. doi: 10.1073/pnas.1817074116
- Negri, R. (2015). Polyacetylenes from terrestrial plants and fungi: recent phytochemical and biological advances. *Fitoterapia* 106, 92–109. doi: 10.1016/j.fitote.2015.08.011
- Ning, J., Moghe, G. D., Leong, B., Kim, J., Ofner, I., Wang, Z., et al. (2015). A feedback-insensitive Isopropylmalate synthase affects acylsugar composition in cultivated and wild tomato. *Plant Physiol.* 169, 1821–1835. doi: 10.1104/pp.15.00474
- Ohlrogge, J., Thrower, N., Mhaske, V., Stymne, S., Baxter, M., Yang, W., et al. (2018). PlantFAdB: a resource for exploring hundreds of plant fatty acid structures synthesized by thousands of plants and their phylogenetic relationships. *Plant J.* 96, 1299–1308. doi: 10.1111/tpj.14102
- Pagani, M., Zachos, J. C., Freeman, K. H., Tipple, B., and Bohaty, S. (2005). Marked decline in atmospheric carbon dioxide concentrations during the Paleogene. *Science* 309, 600–603. doi: 10.1126/science.1110063
- Pichersky, E., and Lewinsohn, E. (2011). Convergent evolution in plant specialized metabolism. *Annu. Rev. Plant Biol.* 62, 549–566. doi: 10.1146/annurev-arplant-042110-103814
- Polturak, G., and Aharoni, A. (2018). “La Vie en Rose”: biosynthesis, sources, and applications of betalain pigments. *Mol. Plant* 11, 7–22. doi: 10.1016/j.molp.2017.10.008
- Qian, Y., Lynch, J. H., Guo, L., Rhodes, D., Morgan, J. A., and Dudareva, N. (2019). Completion of the cytosolic post-chorismate phenylalanine biosynthetic pathway in plants. *Nat. Commun.* 10:15. doi: 10.1038/s41467-018-07969-7962
- Reeves, G., Grangé-Guermente, M. J., and Hibberd, J. M. (2017). Regulatory gateways for cell-specific gene expression in C4 leaves with Kranz anatomy. *J. Exp. Bot.* 68, 107–116. doi: 10.1093/jxb/erw438
- Rippert, P., and Matringe, M. (2002a). Molecular and biochemical characterization of an *Arabidopsis thaliana* aroenate dehydrogenase with two highly similar and active protein domains. *Plant Mol. Biol.* 48, 361–368.
- Rippert, P., and Matringe, M. (2002b). Purification and kinetic analysis of the two recombinant aroenate dehydrogenase isoforms of *Arabidopsis thaliana*. *Eur. J. Biochem.* 269, 4753–4761. doi: 10.1046/j.1432-1033.2002.03172.x
- Rippert, P., Puyaubert, J., Grisolle, D., Derrier, L., and Matringe, M. (2009). Tyrosine and phenylalanine are synthesized within the plastids in *Arabidopsis*. *Plant Physiol.* 149, 1251–1260. doi: 10.1104/pp.108.130070
- Rodríguez-Concepción, M., and Boronat, A. (2015). Breaking new ground in the regulation of the early steps of plant isoprenoid biosynthesis. *Curr. Opin. Plant Biol.* 25, 17–22. doi: 10.1016/j.pbi.2015.04.001

- Romero, R. M., Roberts, M. F., and Phillipson, J. D. (1995). Anthranilate synthase in microorganisms and plants. *Phytochemistry* 39, 263–276. doi: 10.1016/0031-9422(95)00010-15
- Rubin, J. L., and Jensen, R. A. (1979). Enzymology of L-tyrosine biosynthesis in mung bean (*Vigna radiata* [L.] Wilczek). *Plant Physiol.* 64, 727–734. doi: 10.1104/pp.64.5.727
- Sage, R. F., Christin, P.-A., and Edwards, E. J. (2011). The C4 plant lineages of planet earth. *J. Exp. Bot.* 62, 3155–3169. doi: 10.1093/jxb/err048
- Sage, R. F., Khoshravesht, R., and Sage, T. L. (2014). From proto-Kranz to C4 Kranz: building the bridge to C4 photosynthesis. *J. Exp. Bot.* 65, 3341–3356. doi: 10.1093/jxb/eru180
- Sage, R. F., Sage, T. L., and Kocacinar, F. (2012). Photorespiration and the evolution of C4 photosynthesis. *Annu. Rev. Plant Biol.* 63, 19–47. doi: 10.1146/annurev-arplant-042811-105511
- Sage, R. F., and Stata, M. (2015). Photosynthetic diversity meets biodiversity: the C4 plant example. *J. Plant Physiol.* 172, 104–119. doi: 10.1016/j.jplph.2014.07.024
- Sage, T. L., Busch, F. A., Johnson, D. C., Friesen, P. C., Stinson, C. R., Stata, M., et al. (2013). Initial events during the evolution of C4 photosynthesis in 3 species of *Flaveria*. *Plant Physiol.* 163, 1266–1276. doi: 10.1104/pp.113.221119
- Schenck, C. A., Chen, S., Siehl, D. L., and Maeda, H. A. (2015). Non-plastidic, tyrosine-insensitive prephenate dehydrogenases from legumes. *Nat. Chem. Biol.* 11, 52–57. doi: 10.1038/nchembio.1693
- Schenck, C. A., Holland, C. K., Schneider, M. R., Men, Y., Lee, S. G., Jez, J. M., et al. (2017a). Molecular basis of the evolution of alternative tyrosine biosynthetic routes in plants. *Nat. Chem. Biol.* 13, 1029–1035. doi: 10.1038/nchembio.2414
- Schenck, C. A., Men, Y., and Maeda, H. A. (2017b). Conserved molecular mechanism of TyrA dehydrogenase substrate specificity underlying alternative tyrosine biosynthetic pathways in plants and microbes. *Front. Mol. Biosci.* 4:73. doi: 10.3389/fmolb.2017.00073
- Schenck, C. A., and Maeda, H. A. (2018). Tyrosine biosynthesis, metabolism, and catabolism in plants. *Phytochemistry* 149, 82–102. doi: 10.1016/j.phytochem.2018.02.003
- Schillmiller, A., Shi, F., Kim, J., Charbonneau, A. L., Holmes, D., Daniel Jones, A., et al. (2010). Mass spectrometry screening reveals widespread diversity in trichome specialized metabolites of tomato chromosomal substitution lines. *Plant J.* 62, 391–403. doi: 10.1111/j.1365-3113.2010.04154.x
- Schillmiller, A. L., Moghe, G. D., Fan, P., Ghosh, B., Ning, J., Jones, A. D., et al. (2015). Functionally divergent alleles and duplicated Loci encoding an acyltransferase contribute to acylsugar metabolite diversity in *Solanum* trichomes. *Plant Cell* 27, 1002–1017. doi: 10.1105/tpc.15.00087
- Schlüter, U., and Weber, A. P. M. (2016). The road to C4 photosynthesis: evolution of a complex trait via intermediary states. *Plant Cell Physiol.* 57, 881–889. doi: 10.1093/pcp/pcw009
- Schulze, S., Mallmann, J., Burscheidt, J., Koczor, M., Streubel, M., Bauwe, H., et al. (2013). Evolution of C4 photosynthesis in the genus *Flaveria*: establishment of a photorespiratory CO₂ pump. *Plant Cell* 25, 2522–2535. doi: 10.1105/tpc.113.114520
- Siehl, D. L., and Conn, E. E. (1988). Kinetic and regulatory properties of arogenate dehydratase in seedlings of *Sorghum bicolor* (L.) Moench. *Arch. Biochem. Biophys.* 260, 822–829. doi: 10.1016/0003-9861(88)90513-9
- Simpson, J. P., and Ohlrogge, J. B. (2016). A novel pathway for triacylglycerol biosynthesis is responsible for the accumulation of massive quantities of glycerolipids in the surface wax of bayberry (*Myrica pensylvanica*) fruit. *Plant Cell* 28, 248–264. doi: 10.1105/tpc.15.00900
- Simpson, J. P., Thrower, N., and Ohlrogge, J. B. (2016). How did nature engineer the highest surface lipid accumulation among plants? Exceptional expression of acyl-lipid-associated genes for the assembly of extracellular triacylglycerol by Bayberry (*Myrica pensylvanica*) fruits. *Biochim. Biophys. Acta* 1861, 1243–1252. doi: 10.1016/j.bbalip.2016.01.022
- Song, H. S., Brotherton, J. E., Gonzales, R. A., and Widholm, J. M. (1998). Tissue culture-specific expression of a naturally occurring tobacco feedback-insensitive anthranilate synthase. *Plant Physiol.* 117, 533–543. doi: 10.1104/pp.117.2.533
- Sulpice, R., and McKeown, P. C. (2015). Moving toward a comprehensive map of central plant metabolism. *Annu. Rev. Plant Biol.* 66, 187–210. doi: 10.1146/annurev-arplant-043014-114720
- Tanaka, Y., Sasaki, N., and Ohmiya, A. (2008). Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. *Plant J.* 54, 733–749. doi: 10.1111/j.1365-3113.2008.03447.x
- Tholl, D. (2015). Biosynthesis and biological functions of terpenoids in plants. *Adv. Biochem. Eng. Biotechnol.* 148, 63–106. doi: 10.1007/10_2014_295
- Tohge, T., Watanabe, M., Hoefgen, R., and Fernie, A. R. (2013). The evolution of phenylpropanoid metabolism in the green lineage. *Crit. Rev. Biochem. Mol. Biol.* 48, 123–152. doi: 10.3109/10409238.2012.758083
- Tzin, V., and Galili, G. (2010). New insights into the shikimate and aromatic amino acids biosynthesis pathways in plants. *Mol. Plant* 3, 956–972. doi: 10.1093/mp/ssq048
- van de Loo, F. J., Broun, P., Turner, S., and Somerville, C. (1995). An oleate 12-hydroxylase from *Ricinus communis* L. is a fatty acyl desaturase homolog. *Proc. Natl. Acad. Sci. U.S.A.* 92, 6743–6747. doi: 10.1073/pnas.92.15.6743
- Vogt, T. (2010). Phenylpropanoid biosynthesis. *Mol. Plant* 3, 2–20. doi: 10.1093/mp/ssp106
- Vranová, E., Coman, D., and Griseem, W. (2013). Network analysis of the MVA and MEP pathways for isoprenoid synthesis. *Annu. Rev. Plant Biol.* 64, 665–700. doi: 10.1146/annurev-arplant-050312-120116
- Wang, M., Toda, K., and Maeda, H. A. (2016). Biochemical properties and subcellular localization of tyrosine aminotransferases in *Arabidopsis thaliana*. *Phytochemistry* 132, 16–25. doi: 10.1016/j.phytochem.2016.09.007
- Weng, J.-K. (2014). The evolutionary paths towards complexity: a metabolic perspective. *New Phytol.* 201, 1141–1149. doi: 10.1111/nph.12416
- Weng, J.-K., Philippe, R. N., and Noel, J. P. (2012). The rise of chemodiversity in plants. *Science* 336, 1667–1670. doi: 10.1126/science.1217411
- Wichers, H. J., Visser, J. F., Huizing, H. J., and Pras, N. (1993). Occurrence of L-DOPA and dopamine in plants and cell cultures of *Mucuna pruriens* and effects of 2,4-d and NaCl on these compounds. *Plant Cell Tiss. Organ. Cult.* 33, 259–264. doi: 10.1007/BF02319010
- Williams, B. P., Aubry, S., and Hibberd, J. M. (2012). Molecular evolution of genes recruited into C₄ photosynthesis. *Trends Plant Sci.* 17, 213–220. doi: 10.1016/j.tplants.2012.01.008
- Yoo, H., Widhalm, J. R., Qian, Y., Maeda, H., Cooper, B. R., Jannasch, A. S., et al. (2013). An alternative pathway contributes to phenylalanine biosynthesis in plants via a cytosolic tyrosine:phenylpyruvate aminotransferase. *Nat. Commun.* 4:2833. doi: 10.1038/ncomms3833

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Maeda. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Identification and Functional Characterization of Genes Involved in the Biosynthesis of Caffeoylquinic Acids in Sunflower (*Helianthus annuus* L.)

Ketthida Cheevarunapakul¹, Gholamreza Khaksar¹, Pawinee Panpetch¹, Patwira Boonjing¹ and Supaart Sirikantaramas^{1,2*}

OPEN ACCESS

Edited by:

Danièle Werck,
Centre National de la Recherche
Scientifique (CNRS), France

Reviewed by:

David Gagneul,
Lille University of Science and
Technology, France
Alain Hehn,
Université de Lorraine, France

*Correspondence:

Supaart Sirikantaramas
supaart.s@chula.ac.th

Specialty section:

This article was submitted to
Plant Metabolism and
Chemodiversity,
a section of the journal
Frontiers in Plant Science

Received: 01 March 2019

Accepted: 10 July 2019

Published: 31 July 2019

Citation:

Cheevarunapakul K, Khaksar G,
Panpetch P, Boonjing P and
Sirikantaramas S (2019) Identification
and Functional Characterization of
Genes Involved in the Biosynthesis of
Caffeoylquinic Acids in
Sunflower (*Helianthus annuus* L.).
Front. Plant Sci. 10:968.
doi: 10.3389/fpls.2019.00968

¹Molecular Crop Research Unit, Department of Biochemistry, Faculty of Science, Chulalongkorn University, Bangkok, Thailand, ²Molecular Sensory Science Center, Faculty of Science, Chulalongkorn University, Bangkok, Thailand

Sunflower (*Helianthus annuus* L.) sprouts accumulate high amounts of caffeoylquinic acids (CQAs) including chlorogenic acid (5-CQA) and 1,5-diCQA. These compounds, which can be found in many plants, including tomato, globe artichoke, and chicory, have many health benefits, including antioxidant, antihepatotoxic, and antiglycative activities. However, CQA profiles and biosynthesis have not previously been studied in sunflower sprouts. In the present study, we found that 5-CQA and 1,5-diCQA were the major CQAs found in sunflower sprouts. We also identified minor accumulation of other CQAs, namely 3-CQA, 4-CQA, 3,4-diCQA, and 4,5-diCQA. According to genome-wide identification and phylogenetic analysis of genes involved in CQA biosynthesis in sunflower, three genes (*HaHQT1*, *HaHQT2*, and *HaHQT3*) encoding hydroxycinnamoyl CoA:quinic acid hydroxycinnamoyl transferase (HQT) and two genes (*HaHCT1* and *HaHCT2*) encoding hydroxycinnamoyl CoA:shikimate/quinic acid hydroxycinnamoyl transferase (HCT) were identified. Expression analysis of these five genes in hypocotyls and cotyledons strongly suggested that *HaHQT2* could be the main enzyme responsible for CQA biosynthesis, as *HaHQT2* had the highest expression levels. In addition, when transiently expressed in the leaves of *Nicotiana benthamiana*, all three *HaHQTs*, which were soluble and not membrane-bound enzymes, could increase the content of 5-CQA by up to 94% compared to that in a control. Overall, our results increase understanding of CQA biosynthesis in sunflower sprouts and could be exploited by plant breeders to enhance accumulation of health-promoting CQAs in these plants.

Keywords: hydroxycinnamoyl-coenzyme A:quinic acid hydroxycinnamoyl transferase, hydroxycinnamoyl-coenzyme A:shikimate/quinic acid hydroxycinnamoyl transferase, caffeoylquinic acid, sunflower (*Helianthus annuus* L.), sprout, functional characterization

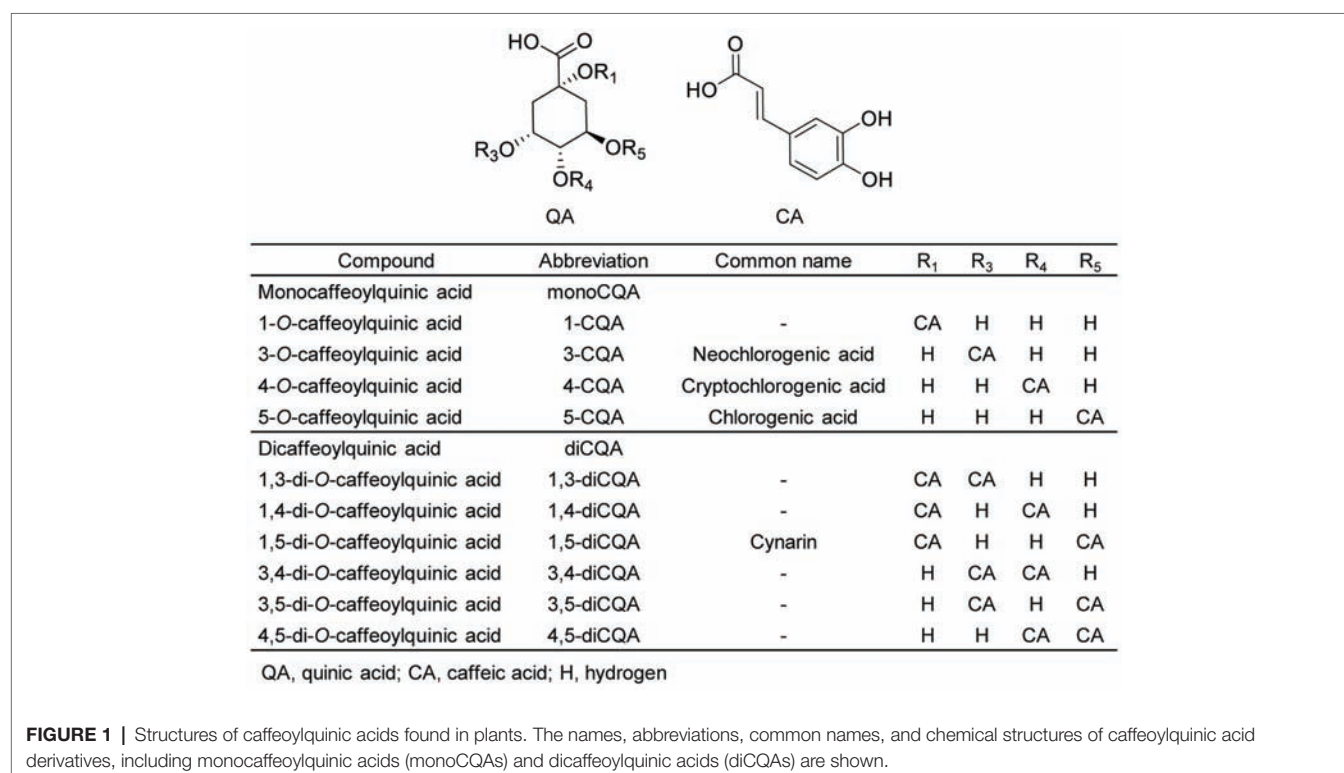
INTRODUCTION

Increased health consciousness among consumers and concerns about the negative health effects of chemical preservatives used in the food industry has led to increase an interest in natural and herbal substances. Fruits and vegetables accumulate a wide range of bioactive compounds with many health-promoting benefits. Among these bioactive compounds, phenolics are of high importance, and caffeoylquinic acids (CQAs) comprise one of the most common phenolic groups. When caffeoyl moieties combine with quinic acid, CQAs are formed. These CQAs can be categorized into various groups based on the position, number, and identity of their acyl group. The moncaffeoylquinic acid (monoCQA) group includes 1-CQA, 3-CQA (known as neochlorogenic acid), 4-CQA (known as cryptochlorogenic acid), and 5-CQA (known as chlorogenic acid). The dicaffeoylquinic acid (diCQA) group includes 1,3-diCQA, 1,4-diCQA, 1,5-diCQA, 3,4-diCQA, 3,5-diCQA, and 4,5-diCQA (Figure 1).

CQAs can be found in numerous plant species, including *Cichorium intybus* (chicory; Legrand et al., 2016), *Cynara cardunculus* L. var. *scolymus* (globe artichoke; Moglia et al., 2016), and *Helianthus annuus* L. (sunflower; Sun et al., 2012) in the family Asteraceae; *Solanum lycopersicum* L. (tomato) and *Nicotiana tabacum* L. (tobacco) in the family Solanaceae (Niggeweg et al., 2004); *Coffea* spp. (coffee; Lallemand et al., 2012) in the family Rubiaceae; and *Ipomoea batatas* L. (sweet potato; Kojima and Kondo, 1985) and *Ipomoea aquatica* (water spinach; Lawal et al., 2016) in the family Convolvulaceae. The antioxidant, anti-inflammatory, anti-hypertension, and

antimicrobial properties of CQAs are well documented by previous studies. In addition, several *in vitro* and *in vivo* studies have shown additional benefits of CQAs, such as a reduction in the risk of cardiovascular diseases (de Sotillo and Hadley, 2002), hepatoprotective properties (Salomone et al., 2017), and inhibition of HIV replication and integration (McDougall et al., 1998; Gu et al., 2007). Notably, CQAs can confer resistance to abiotic stressors such as UV light (Cle et al., 2008) and to biotic stressors (Niggeweg et al., 2004; Leiss et al., 2009; Legrand et al., 2016) in plants. The bioactivity of CQAs mainly depends on their isomerization. For example, the number and position of caffeic acid moieties in diCQAs affect their antioxidant properties (Xu et al., 2012).

Sunflower (*Helianthus annuus* L.) seeds and sprouts are rich in phenolic compounds and vitamins and thus exhibit a wide variety of potential health-beneficial characteristics, including anti-inflammatory, antimicrobial, antioxidant, antihypertensive, and wound-healing properties (Fowler, 2006; Bashir et al., 2015; Guo et al., 2017). In a study by Weisz et al. (2009), monoCQA and diCQA contents comprised up to ~3,359 and 460 mg per 100 g dry weight of seed kernels, respectively. Among the 11 phenolic compounds analyzed in sunflower seed kernels, 5-CQA was the most abundant. In another study by Sun et al. (2012), the antiglycative and antioxidant characteristics of four edible sprouts were investigated, and it was found that sunflower sprout extract exhibited similar antiglycative properties compared with aminoguanidine, a well-known synthetic antiglycative agent. The strong antioxidant and antiglycative properties of sunflower sprouts were attributed to their rich 1,5-diCQA content. Pająk et al. (2014) examined the effect of germination



on nutritional value (total phenolics and flavonoids) and antioxidant properties of sunflower seeds. Interestingly, they found that total phenolics, flavonoids, and antioxidant capabilities were significantly higher in sunflower sprouts than seeds. Moreover, HPLC profiling of sunflower phenolics revealed that CQA content was 3.7-fold higher in sprouts than seeds. Taken together, these observations prompted us to investigate sunflower sprouts for our study.

CQAs are biosynthesized *via* the phenylpropanoid pathway in plants (Comino et al., 2009). The starting point of this pathway is the aromatic amino acid phenylalanine (Phe), which is deaminated by phenylalanine ammonia lyase (PAL) to form cinnamic acid. Then, cinnamate-4-hydroxylase (C4H) and 4-coumarate coenzyme A ligase (4CL) sequentially

convert cinnamic acid to form *p*-coumaroyl-CoA. Two possible routes have been proposed for the next step in CQA synthesis. In the first route, hydroxycinnamoyl-CoA:quinic acid hydroxycinnamoyl transferase (HQT) converts *p*-coumaroyl-CoA to coumaroylquinic acid, which is then hydroxylated by *p*-coumarate-3'-hydroxylase (C3'H) to form CQA (Niggeweg et al., 2004; Comino et al., 2009; Menin et al., 2010). In the alternative second route, hydroxycinnamoyl-CoA:shikimate/quinic acid hydroxycinnamoyl transferase (HCT) catalyzes the formation of *p*-coumaroylshikimate from *p*-coumaroyl-CoA. The *p*-coumaroylshikimate is subsequently hydroxylated to caffeoylshikimic acid by C3'H (Mahesh et al., 2007; Moglia et al., 2009). Notably, HQT and HCT both catalyze reversible reactions (Figure 2).

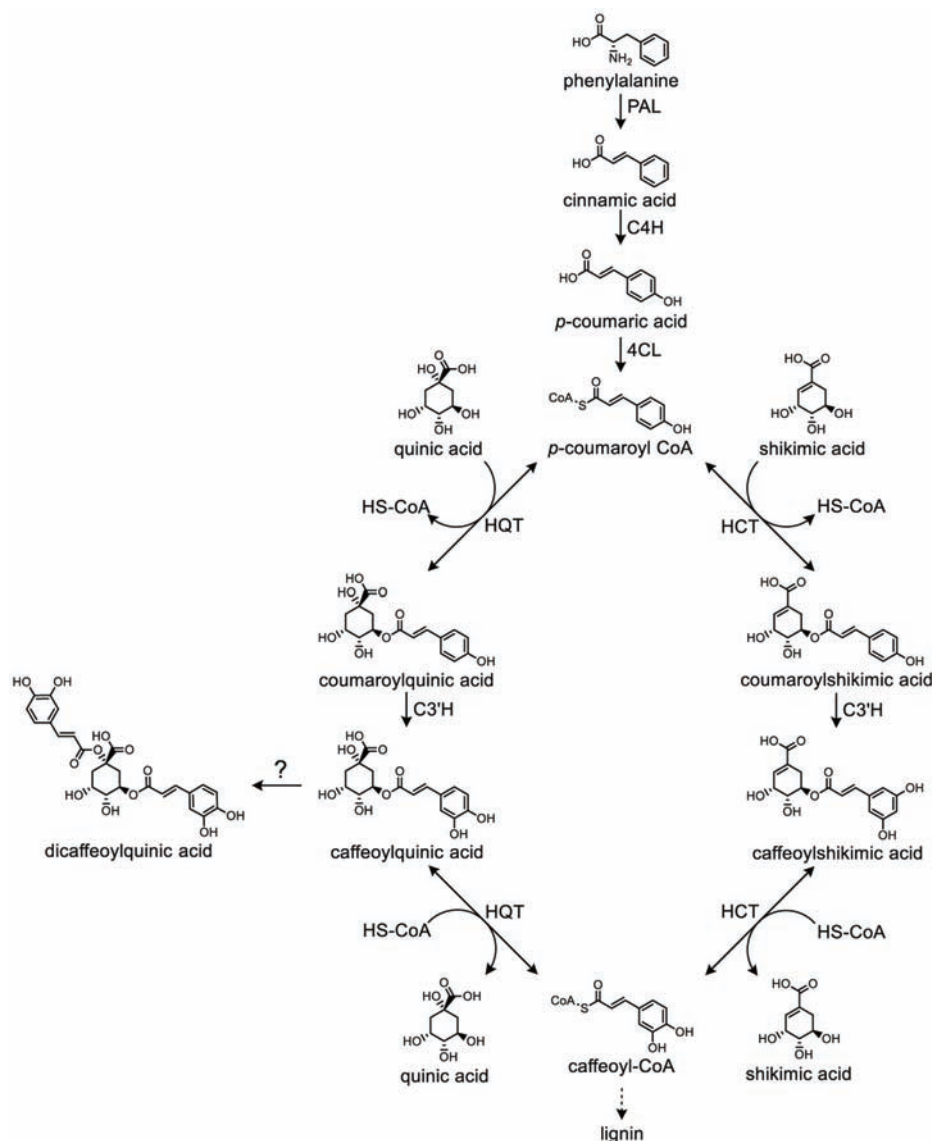


FIGURE 2 | Biosynthetic pathway of caffeoylquinic acids in plants. The roles of phenylalanine ammonia lyase (PAL), cinnamate 4-hydroxylase (C4H), 4-hydroxycinnamoyl-CoA ligase (4CL), hydroxycinnamoyl-CoA:quinic acid hydroxycinnamoyl transferase (HQT), hydroxycinnamoyl-CoA:shikimate/quinic acid hydroxycinnamoyl transferase (HCT), and *p*-coumaroyl ester 3'-hydroxylase (C3'H) are depicted.

HQT and HCT enzymes belong to the BAHD superfamily of plant-specific acyl-CoA-dependent acyltransferases (Yu et al., 2009; Tuominen et al., 2011). However, our knowledge regarding the biosynthesis of diCQAs is still limited. In a study by Lallemand et al. (2012), a recombinant HCT enzyme cloned from coffee was shown to synthesize diCQAs from 5-CQA. In addition, the enzyme HQT was reported to convert 5-CQA to diCQAs in tomato (Moglia et al., 2014). Legrand et al. (2016) identified two HCTs (HCT1 and HCT2) and three HQTs (HQT1, HQT2, and HQT3) in chicory. Notably, increased levels of 3-CQA were detected in *N. benthamiana* leaves transiently expressing HQT1 or HCT1. Moreover, genes involved in CQA biosynthesis, including *HQT*, *HCT*, *C3'H*, *Acyltransf_1*, and *Acyltransf_2*, have been isolated and characterized in artichoke (Comino et al., 2007, 2009; Moglia et al., 2009; Menin et al., 2010). Moglia et al. (2016) found two *HCT*s and three *HQT*s in artichoke, which is the same number of *HCT*s and *HQT*s found in chicory. To the best of our knowledge, there are currently no published studies in the literature on how CQAs are formed during sprouting stages and how high the CQA content is. Therefore, we focused on sunflower sprout, which is currently gaining popularity among health-conscious consumers. In addition, none of *HQT*s and *HCT*s have been identified in sunflower. Acquiring a deeper understanding of CQA biosynthetic enzymes in sunflower is critically important to aid efforts to biofortify sunflower sprouts as functional foods. In the present study, we report the identification and functional characterization of key genes involved in CQA biosynthesis in sunflower sprouts.

MATERIALS AND METHODS

Plant Material and Growth Conditions

Sunflower (*H. annuus* L.) seeds were purchased from a local supplier ("Super Top," Thailand). The seeds were first washed, soaked in tap water for 8 h, and wrapped with wet cheesecloth overnight. Next, they were germinated on coconut dust at 30°C under 60% relative humidity for 9 days, with dark conditions for the first 48 h, followed by a 12/12 h light/dark photoperiod for the remaining days. The seed coat was gently removed at hour 56 (after exposure to the light for 8 h; **Supplementary Figure S1A**). To compare metabolites and genes expression of the sample obtained from different developmental stages and tissues, the experiment was divided into two sets. For the first set, whole plants excluding roots were collected at different time points (days 3–9, 2 h after light exposure at each day; **Supplementary Figure S1B**) after germination. For the second set, cotyledons and stems were collected similarly at day 5. For both sets, there were five replicates per time point, and each replicate consisted of 10 sprouts. Replicates were collected separately, frozen in liquid nitrogen, ground into a fine powder using an MM400 mixer mill (Retsch®, Germany) at 30 Hz for 1 min, and then stored at –80°C until further use. In addition, half of the samples were also freeze dried for HPLC analysis.

To grow five-week-old plants for an agroinfiltration experiment, *N. benthamiana* seeds were sown on peat moss

and grown in a controlled-climate room at 25°C and with a 16/8 h light/dark photoperiod (artificial light of 4,500 Lux). Two-week-old plants were transplanted individually into pots and were left to continue growing under the same conditions.

Determination of Caffeoylquinic Acids Contents of Plant Tissues

To analyze CQA contents of plant tissues, 20 mg dry weight of sunflower sprout tissue and 20 mg fresh weight of *N. benthamiana* leaves were extracted with 1 ml of 80% (v/v) methanol containing an internal standard, 0.05 g L^{–1} puerarin. The reactions were mixed vigorously at 15°C for 15 min by shaking at 1,500 rpm and then centrifuged at 12,000 × g for 15 min. Supernatant was collected and filtered through 0.2 µm nylon syringe filters.

A Shimadzu UFLC system equipped with an SPD-M20A photodiode array detector (Shimadzu, Japan) and Kinetex® C18 (250 mm × 4.6 mm, 5 µm; Phenomenex®, USA) was used to analyze 10 µl of the extract from sunflower sprouts and *N. benthamiana* leaves. Chromatographic separation was performed using 0.1% (v/v) TFA in water (solvent A) and 0.1% (v/v) TFA in acetonitrile (solvent B) as the mobile phase. The following elution gradient was used: 5% B for 5 min, 5–15% B for 10 min, a 25-min hold, 15–100% B for 4 min, a 2-min hold, 100–5% B for 4 min, and a 5-min hold. The flow rate was set at 1.5 ml min^{–1}, and the column oven temperature was maintained at 40°C. UV spectra were acquired in the range of 190–800 nm, and chromatograms were obtained at 320 nm. Peaks corresponding with the retention time and UV spectrum of a commercial standard were identified as CQAs. Amounts of each CQA were calculated according to the calibration curve in the range of 0.5–0.007825 mg ml^{–1}. Puerarin was used as an internal standard (Sigma-Aldrich, USA). All CQA standards used in this study were purchased from Carbosynth, England.

Additionally, to confirm identities of the CQAs, the components were analyzed using an Agilent UHPLC system (Agilent Technologies, USA) using Kinetex® C18 (250 mm × 4.6 mm, 5 µm; Phenomenex®, USA). The following elution gradient was used: 0–5% B for 5 min, 5–15% B for 30 min, a 65-min hold, 15–100% B for 5 min, a 5-min hold, 100–5% B for 5 min, and a 10-min hold. The flow rate was set at 0.5 ml min^{–1}, and the column oven temperature was maintained at 40°C. For MS/MS analysis, QTRAP® 4,500 MS/MS System (AB Sciex™, USA) in multiple reaction monitoring (MRM) and negative ionization mode (ESI[–]) was used. Operating conditions for MS analysis were as follows: heat block temperature of 500°C, curtain nitrogen gas 30 psi, nebulizer and auxiliary gases of 50 psi, collision nitrogen gas at medium position, ionization voltage of –4,500 V, and entrance potential (EP) of –10. For the tested compounds, the following transition under optimal instrumental conditions of collision energy (CE) of –35 eV, declustering potential (DP) of –50 V, and collision cell exit potential (CXP) of –12 V.

Identification of Putative *HQT* and *HCT* Genes in Sunflower (*HaHQTs* and *HaHCTs*)

The *HQT* of tomato (*Solanum lycopersicum*; NP_001234850.2; Moglia et al., 2014) was used as a query for tBlastn search against the sunflower genome database HA412HO bronze assembly¹. The open reading frames of *HQT* and *HCT* were identified. Then, using EMBL-EBI Clustal Omega (McWilliam et al., 2013), the amino acid sequences of putative *HaHQTs* and *HaHCTs* were aligned with well-characterized *HQTs/HCTs* belonging to different plant species (**Supplementary Figure S2**). These candidates including *HaHQT1* (accession number MK598073), *HaHQT2* (accession number MK598074), *HaHQT3* (accession number MK598075), *HaHCT1* (accession number MK598076), and *HaHCT2* (accession number MK598077) were selected for further study.

Phylogenetic Analysis

Amino acid sequences of putative *HaHQTs* and *HaHCTs* were aligned with sequences of previously characterized enzymes using BioEdit ClustalW multiple alignment (Hall, 1999), and a Neighbor Joining (NJ) tree was created using MEGA7 software (Kumar et al., 2016) with 1,000 bootstrapped data sets.

RNA Isolation, cDNA Synthesis, and Cloning of Putative *HaHQTs*

Total RNA was isolated from 100 mg fresh weight of sunflower sprouts using TRI reagent® (Molecular Research Center, Inc., USA). Next, RNA concentration and integrity were analyzed by measuring A_{260} and A_{280} on an Eppendorf Biophotometer® D30 (Eppendorf, Germany) and by agarose gel electrophoresis, respectively. The RNA was treated with RNase-free DNase I (Thermo Fisher Scientific, USA), and then, the first strand cDNA was synthesized by RevertAid Reverse Transcriptase using oligo(dT)₂₀ primers (Thermo Fisher Scientific, USA) according to the manufacturer's instructions.

The full-length putative *HaHQTs* were amplified with Phusion Hot Start II High-Fidelity DNA Polymerase (Thermo Fisher Scientific, USA) using the prepared cDNA of sunflower sprout as a template. Then, the amplified DNA was cloned into pCR™8/GW/TOPO®TA vectors (Invitrogen, USA) resulting in pCR™8/GW/TOPO®-*HaHQTs* and subsequently sequenced. One clone of each putative gene was used for further study (sections Promoter Analysis and Transient Overexpression of *HaHQTs* in *N. benthamiana*).

Gene Expression Analysis of Sunflower Sprout

Total RNA was extracted from sunflower sprouts as described above (section RNA Isolation, cDNA Synthesis, and Cloning of Putative *HaHQTs*). Then, qRT-PCR was performed using gene-specific primers (**Supplementary Table S1**). Eukaryotic translation initiation factor 5A (*ETIF5A*; XM_022156448.1), elongation factor 2 (*EF2*; XM_022137686.1), and actin 7 (*ACT7*;

XM_022154554.1) of sunflower were used as reference genes (Ochogavía et al., 2017). Reactions were conducted in volumes of 10 µl in a 96-well PCR plate using Luna® universal qPCR master mix (New England Biolabs®, USA). A CFX Connect™ Real-Time PCR Detection System and CFX Manager™ Software (BIO-RAD, USA) were used to conduct PCR, and melting curve analysis was used to confirm the existence of a single product. Relative expression level of each gene was calculated using $2^{-\Delta C_t}$ (Schmittgen and Livak, 2008) according to the average C_t values of three reference genes (Beekman et al., 2011).

For droplet digital PCR (ddPCR), a 20-µl reaction mixture containing gene-specific primers (**Supplementary Table S1**), QX200™ ddPCR™ EvaGreen Supermix (BIO-RAD, USA), and cDNA was generated as a droplet with QX200™ Droplet Generation Oil for EvaGreen (BIO-RAD, USA) using QX200™ droplet generator (BIO-RAD, USA). EF2 was used as a reference gene. The PCRs were performed in a 96-well PCR plate using a T100™ Thermal Cycler (BIO-RAD, USA). After amplification, QX200™ Droplet Reader (BIO-RAD, USA) was used to measure the fluorescence intensity of each individual droplet. Absolute transcript levels (copies/20 µl reaction) were processed using QuantaSoft™ Software (BIO-RAD, USA). Relative transcript number of each gene was presented as a ratio of the absolute transcript levels (copies/20 µl reaction) of the target gene to the reference gene (Taylor et al., 2017).

Promoter Analysis

The 2,000 bp upstream regions of start codon of putative *HaHQTs* were *in silico* scanned for regulatory elements using MatInspector (Cartharius et al., 2005). The genomic localization of the analyzed promoters was Chr10: 227227684...227229684 for *HaHQT2* and Chr2: 166074069...166076069 for *HaHQT3*.

Transient Overexpression of *HaHQTs* in *Nicotiana benthamiana*

The putative *HaHQTs* from pCR™8/GW/TOPO®-*HaHQTs* were transferred into pEAQ-HT-DEST1 (pEAQ1) expression vectors (Peyret and Lomonosoff, 2013) using Gateway® LR Clonase® II (Invitrogen, USA). The resultant pEAQ1-*HaHQTs* were then transformed into *Agrobacterium tumefaciens* LBA4404 by electroporation.

A. tumefaciens colonies containing each construct were grown in 25 ml of LB broth containing 50 mg L⁻¹ kanamycin, 50 mg L⁻¹ streptomycin, and 50 mg L⁻¹ rifampicin and shaken at 250 rpm at a temperature of 30°C overnight. Cells were harvested by centrifugation at 3,000 × g for 10 min and washed in MM buffer twice (10 mM MES and 10 mM MgCl₂, pH 5.6). Then, the pellet was resuspended in MM buffer to an optical density of 0.4 at OD₆₀₀, and acetosyringone was added to a final concentration of 100 mg L⁻¹. The culture solution was incubated at room temperature for 2 h. Genes of interest were transferred into the abaxial leaves of 5-week-old plants, by first nicking the leaf on the backside with a needle and then infiltrating the gene-harboring *A. tumefaciens* using a needleless 1-ml syringe. After 5 days, the infiltrated leaves were collected, frozen in liquid nitrogen, and ground into a fine powder for HPLC analysis.

¹www.sunflowergenome.org

Subcellular Localization

In silico subcellular prediction of localization was performed using the iPSORT (Bannai et al., 2002), WoLF PSORT (Horton et al., 2007), LOCALIZER (Sperschneider et al., 2017), TargetP (Emanuelsson et al., 2000), and ChloroP servers (Emanuelsson et al., 1999).

For *in planta* experiments on subcellular localization, four biological replicates were used. First, *HaHQTs* were amplified with Phusion Hot Start II High-Fidelity DNA Polymerase (Thermo Fisher Scientific, USA) using pCR™8/GW/TOPO®-*HaHQTs* as templates. The primers (excluding stop codons) listed in **Supplementary Table S1** were used in the PCRs. The PCR products were cloned into pCR™8/GW/TOPO®TA vectors (Invitrogen, USA), and nucleotide sequences were verified. Then, the *HaHQTs* were transferred into the C-terminal green fluorescent protein (GFP)-fused destination vector pGWB5 (Nakagawa et al., 2007) using Gateway® LR Clonase® II (Invitrogen, USA), generating pGWB5-*HaHQTs*. The pGWB5-*HaHQTs* were then transformed into *A. tumefaciens* LBA4404 by electroporation.

A. tumefaciens containing each construct and *A. tumefaciens* containing a silencing suppressor *p19* gene (Lindbo, 2007) were co-infiltrated into 5-week-old plants (section Plant Material and Growth Conditions) as in section Transient Overexpression of *HaHQTs* in *N. benthamiana* but with some modifications. In brief, cells obtained from each culture were washed, suspended in MM buffer, and adjusted to an optical density of 0.8 at OD₆₀₀. The culture suspensions of each *A. tumefaciens* harboring pGWB5-*HaHQT* construct were then mixed with that of *A. tumefaciens* harboring *p19* at a ratio of 1:1. Then, acetosyringone was added to a final concentration of 100 mg L⁻¹. At 3 days after infiltration, protein localization was visualized under FluoView® FV10i-DOC confocal laser scanning microscope (Olympus, Japan). Excitation/emission of GFP, autofluorescence of chloroplast, and phase contrast detection were recorded at 473/510, 559/600, and 559/600 nm, respectively.

Statistical Analyses

Statistical analyses were performed using IBM® SPSS® Version 22.0 (IBM, USA) statistical software. Following one-way ANOVA, mean concentrations of CQAs and expression levels of genes were compared between days for each CQA or gene type using Duncan's multiple-range test ($p < 0.05$). In addition, concentration of each CQA and expression levels of genes were compared between hypocotyl and cotyledon tissue types by Student's *t* test ($p < 0.05$).

RESULTS

Caffeoylquinic Acids Profiling in Sunflower Sprouts

Although CQAs in sunflower seeds have been reported (Weisz et al., 2009), previous studies have, to the best of our knowledge, not clearly quantified and characterized CQA content in sunflower sprouts. Therefore, we analyzed CQAs during germination from days 3 to 9. Six CQAs were identified

(Figure 3A), and 1,5-diCQA was the most abundant. Accumulation of 1,5-diCQA increased during germination, reaching a maximum of ~15 mg/g dry weight (Figure 3B). This increasing accumulation level during sprouting was also observed in other CQAs, including 3-CQA, 4-CQA, 3,4-diCQA, and 4,5-diCQA. Notably, the amount of the second most abundant derivative, 5-CQA, did not significantly change over the period. In addition, at day 5 post-germination, we profiled CQA content in two sunflower sprout tissue types: hypocotyl and cotyledon. Cotyledons accumulated much higher levels of 5-CQA and 1,5-diCQA than hypocotyls and contained ~6-fold higher concentrations of 5-CQA. The other CQAs were detected at much lower levels in both tissues. The identities of all CQAs were also confirmed using LC-MS to compare their fragmentation patterns and molecular masses with those of authentic standards (Supplementary Figure S3).

Genome-Wide Identification and Phylogenetic Analysis of *HaHQTs* and *HaHCTs*

A total of five genes encoding sunflower HQTs and HCTs were identified. Multiple alignment of amino acid sequences of all identified HQTs and HCTs showed the conserved motifs of HXXXD and DFGWG, which are signature to the members of the BAHD superfamily (Supplementary Figure S2). Phylogenetic analysis revealed that three *HaHQTs* and two *HaHCTs* were clustered together with HQTs and HCTs from chicory and globe artichoke (Figure 4). The number of HQT and HCT isoforms found in sunflower was identical to the number identified previously in chicory and globe artichoke. We annotated *HaHQTs* (*HaHQT1*, *HaHQT2*, and *HaHQT3*) and *HaHCTs* (*HaHCT1* and *HaHCT2*) based on their clustering with the previously characterized HQTs and HCTs from those two species. Moreover, these *HaHQTs* and *HaHCTs* were located on different chromosomes, e.g., *HaHQT1* on chromosome 9, *HaHQT2* on chromosome 10, *HaHQT3* on chromosome 2, *HaHCT1* on chromosome 16, and *HaHCT2* on chromosome 5.

Gene Expression Profiles of *HaHQTs* and *HaHCTs* in Sunflower Sprouts

To investigate a possible correlation between CQA content and expression levels of corresponding biosynthetic genes, qRT-PCR was used to analyze the expression profiles of identified *HaHQTs* and *HaHCTs*. Sunflower sprouts for this analysis were sampled at seven time points, from days 3 to 9 after germination. As shown in Figure 5A, expression level of *HaHQT1* was not that much different among these time points. However, the expression level of *HaHQT2* was peaked at day 3 and was followed by a significant decrease from day 4. Similarly, expression level of *HaHQT3* was peaked at day 3 and was decreased until day 9, whereas expression level of *HaHCT1* was constant from days 3 to 7 and was significantly decreased from days 7 to 8. Then, it was kept constant again. Noticeably, expression level of *HaHCT2* was increased during germination period. In addition, we investigated the expression levels of these genes in hypocotyls and cotyledons

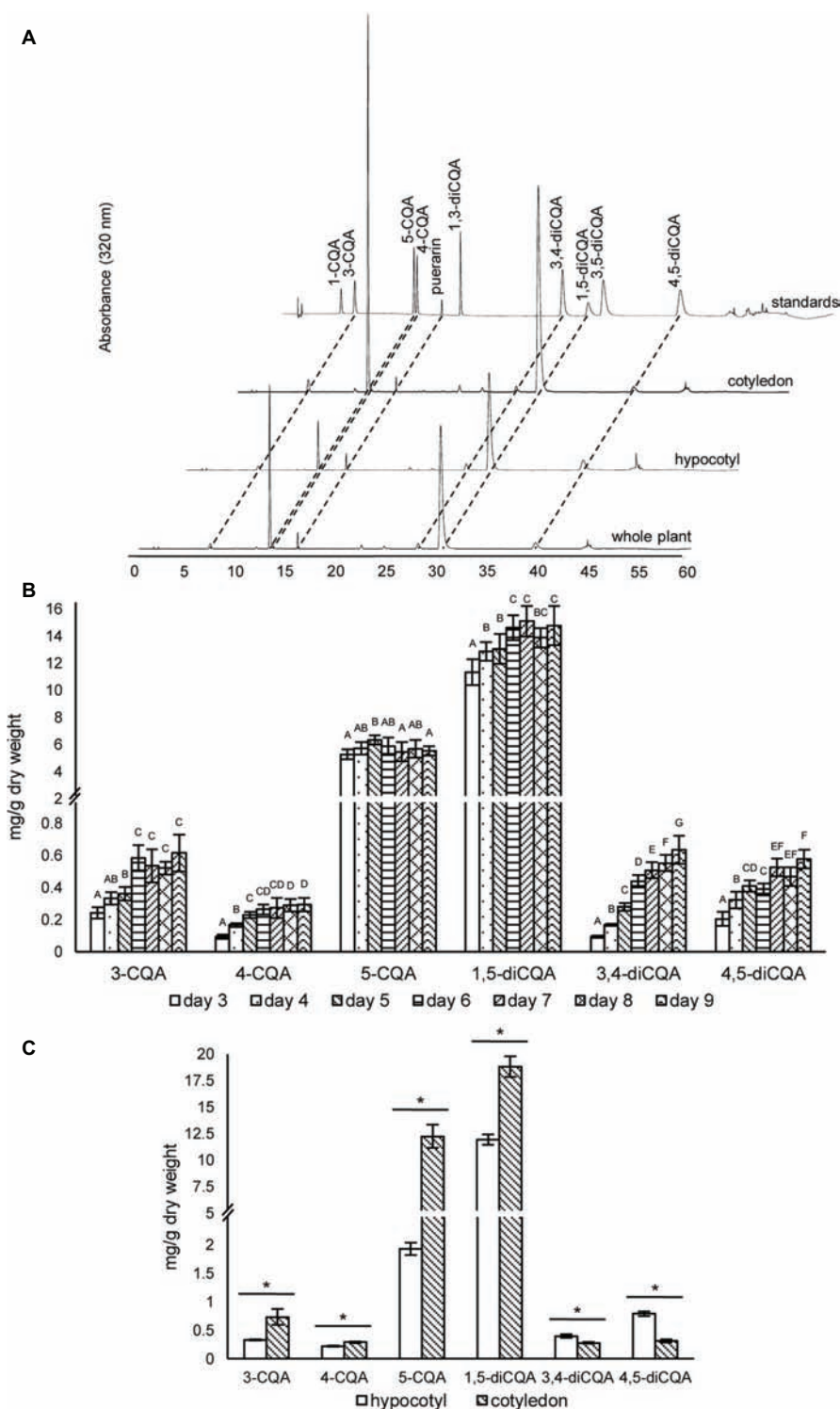


FIGURE 3 | HPLC profiles and concentrations of caffeoylquinic acids in sunflower sprout extracts from different time points during germination and from different tissues. HPLC chromatograms of caffeoylquinic acids of sunflower extracts from whole sprout, hypocotyl, and cotyledon, using puerarin as an internal standard **(A)**; extracts of sunflower sprouts sampled at different time points (days 3–9) during the germination period were measured by HPLC; comparisons are shown for different time points within each caffeoylquinic acid derivative. Bars represent the mean values \pm standard deviation (SD) of five biological and independent replicates; for each derivative, different alphabets indicate significant differences according to Duncan's multiple-range test ($p < 0.05$) **(B)**; and tissue specific concentrations of caffeoylquinic acid derivatives measured in the hypocotyls and cotyledons of sunflower sprouts sampled at 5 day post-germination; an asterisk (*) above the bars indicates a significant difference between the two tissues (Student's t test, $p < 0.05$) **(C)**.

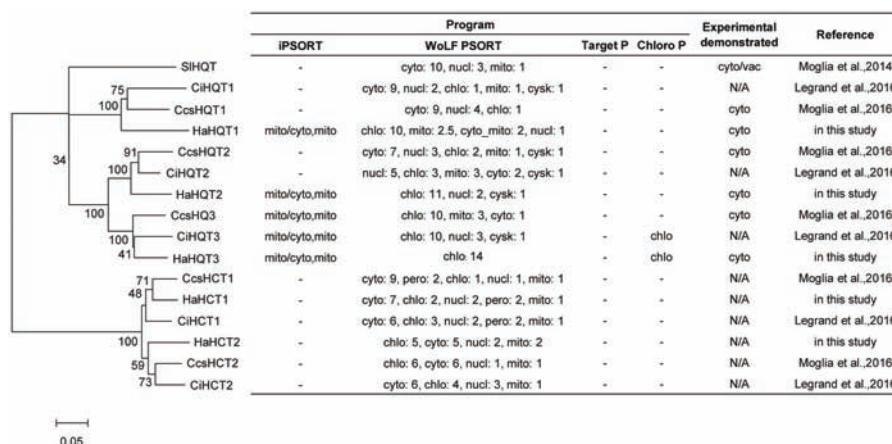


FIGURE 4 | Phylogenetic analysis and *in silico* subcellular localization prediction of sunflower HQTs and HCTs. Sunflower HQTs and HCTs were aligned with HQTs and HCTs from other representative species, including globe artichoke (CcsHQT1; ABK79689.1, CcsHQT2; ADL62854.1, CcsHQT3; ADL62855.1, CcsHCT1; AAZ80046.1, CcsHCT2; KVH99042.1), chicory (CIHQT1; ANN12610.1, CIHQT2; ANN12611.1, CIHQT3; ANN12612.1, CIHCT1; ANN12608.1, CIHCT2; ANN12609.1), and tomato (SIHQT; NP_001234850.2). The phylogenetic tree was constructed with MEGA7 and the neighbor-joining method, using a bootstrap test of phylogeny. *In silico* subcellular localization was predicted using web-based tools, including iPSORT, WoLF PSORT, TargetP, and ChloroP (cyto, cytosol; nucl, nucleus; mito, mitochondria; chlo, chloroplast; cysk, cytoskeleton).

at day 5. Interestingly, *HaHQT1* and *HaHQT2* were expressed at significantly higher levels in cotyledons than in hypocotyls, while that of *HaHQT3* was significantly higher in hypocotyl than in cotyledon (**Figure 5B**). However, the expression levels of two *HaHCTs* were not significantly different between the two tissues. Gene expression analysis by ddPCR confirmed the higher expression level of *HaHQT2* than other *HaHQTs* and *HaHCTs* (**Figure 5C**). These results provided compelling evidence that *HaHQT2* could be the main CQA biosynthetic gene in the sunflower sprouts. Therefore, we selected *HaHQTs* for further functional characterization.

Promoter Analysis

To gain more insights into the regulatory network controlling the expression levels of *HaHQT* genes, we analyzed the promoter regions of those *HaHQT* genes. Due to the incomplete genome database, only promoter regions of *HaHQT2* and *HaHQT3* were analyzed. As shown in **Supplementary Table S2**, these two promoters share some common motifs such as phytochrome, defense response, circadian rhythm, axillary bud outgrowth, light, stress, and phytohormone (auxin, salicylic acid, abscisic acid, jasmonate, gibberellin, ethylene, cytokinin) responsive elements as well as sulfur and sucrose responsive elements. In addition, we found several binding elements for MYB and Dof transcription factors. These two transcription factors were reported to be involved in the phenylpropanoid biosynthesis.

Subcellular Localization of *HaHQTs*

In silico subcellular analysis of sunflower *HaHQTs* predicted their localization in either cytosol or chloroplast (**Figure 4**) with no detection of nuclear localization signal (NLS) in all *HaHQTs*. For assessment *in planta*, *Agrobacteria* harboring each GFP-fused expression construct (pGWB5-*HaHQT1*,

pGWB5-*HaHQT2*, or pGWB5-*HaHQT3*) together with the gene-silencing suppressor *p19* were infiltrated in *N. benthamiana* leaves. Protein localization was analyzed using a confocal laser scanning microscope. *In planta*, all three GFP-tagged *HaHQTs* were soluble and not membrane-bound proteins, possibly localized in the cytosol (**Figure 6; Supplementary Figure S4**). Fluorescence signals were also detected in the nucleus. In addition, observation of mesophyll clearly confirmed that these *HaHQTs* were not localized in chloroplast (**Supplementary Figure S5**).

Transient Expression of *HaHQTs* in *Nicotiana benthamiana*

In *N. benthamiana* leaves infiltrated to transiently express *HaHQTs*, levels of 5-CQA were significantly higher than the control (**Figure 7**). *HaHQT2* and *HaHQT3* also increased the level of 4-CQA. *HaHQT3* is the only isoform that could increase the level of all monoCQAs and 1,3-diCQA. The levels of 1,5-diCQA and 4,5-diCQA did not show any significant differences between the infiltrated leaves versus the control. These results indicated that all *HaHQTs* were involved in CQA biosynthesis.

DISCUSSION

Mono and diCQAs are known to be beneficial to human health. The monoCQA 5-CQA, also known as chlorogenic acid, has a wide variety of health benefits (see review Naveed et al., 2018) and is one of the most abundant CQAs. Although sunflower seed kernels accumulate much higher amounts of monoCQAs than diCQAs (Weisz et al., 2009), we found significantly higher content of diCQAs than of monoCQAs in sunflower sprouts. Our results concur with those of Sun et al.

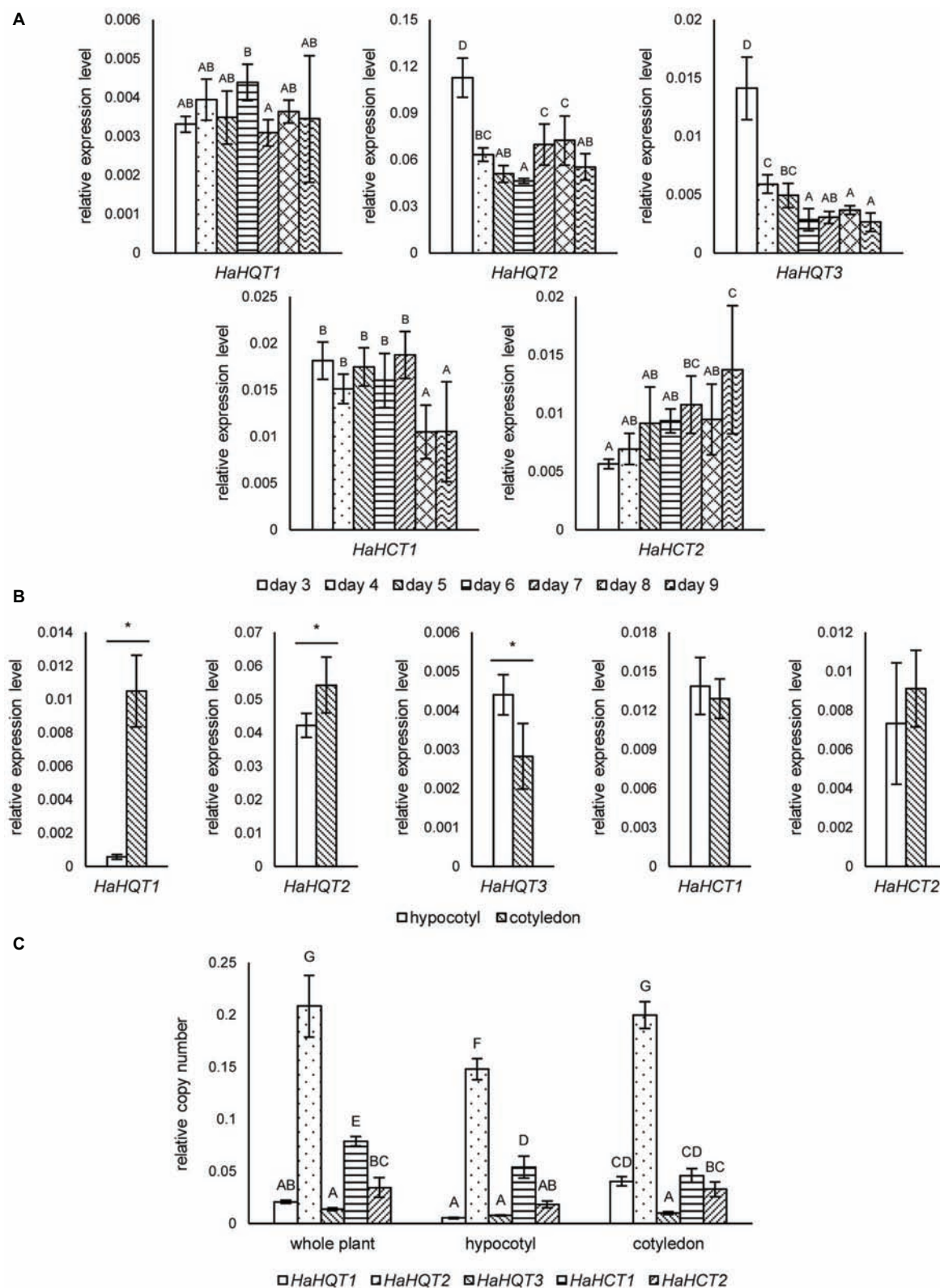


FIGURE 5 | qRT-PCR analysis of sunflower *HQTs* and *HCTs* at different germination time points and tissues. Relative expression levels of sunflower *HQTs* (*HaHQTs*) and *HCTs* (*HaHCTs*) measured at different time points (days 3–9) during the germination period. Bars represent the mean values \pm standard deviation

(Continued)

FIGURE 5 | (SD) of five biologically independent replicates; comparisons are shown among different time points within each gene; different alphabets indicate significant differences according to Duncan's multiple-range test ($p < 0.05$) (A). Tissue specific gene expression analysis of *HaHQTs* and *HaHCTs* sampled at day 5 post-germination; for each gene, comparisons are shown between the hypocotyl and cotyledon; an asterisk (*) above the bars indicates a significant difference between the two tissues (Student's t test, $p < 0.05$) (B). Absolute gene expression analysis of *HaHQTs* and *HaHCTs* from whole plant, hypocotyl, and cotyledon using digital droplet PCR (ddPCR). Bars represent the mean values \pm standard deviation (SD) of five biologically independent replicates; comparisons are shown for all genes among different tissues; different alphabets indicate significant differences according to Duncan's multiple-range test ($p < 0.05$) (C).

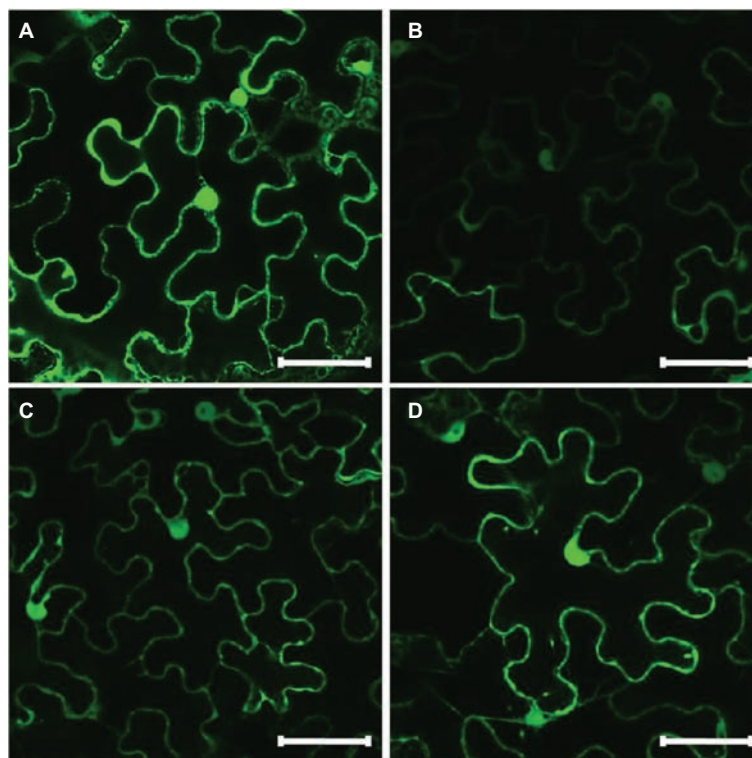


FIGURE 6 | Subcellular localization of GFP-tagged *HaHQTs* in the epidermis of *Nicotiana benthamiana* leaves. Confocal microscopy image of epidermal cells from *N. benthamiana* leaves infiltrated with pGWB2-GFP (control) (A), pGWB5-*HaHQT1* (B), pGWB5-*HaHQT2* (C), or pGWB5-*HaHQT3* (D). Bars = 50 μ m.

(2012) in which 1,5-diCQA is the most abundant CQA in sprouts (Figure 3B). The lower level of monoCQAs found in sprouts could be related to the role of monoCQAs as intermediates used for both diCQA biosynthesis and lignification in growing hypocotyls during sprouting (Escamilla-Treviño et al., 2014). Although 3-CQA, 4-CQA, 1,5-diCQA, 3,4-diCQA, and 4,5-diCQA contents tended to increase in sunflower sprouts until day 9 (Figure 3B), we did not analyze their content after day 9. This was because sunflower sprouts available in markets are generally harvested at day 5 or 6, when the average hypocotyl length is 12–15 cm (Supplementary Figure S1B). Therefore, we designed our experiment to include sampling shortly before (days 3–4) and after (days 7–9) the common harvesting period. Harvest after day 9 would negatively affect the texture of the sprouts, e.g. they would be less crisp, which is inconsistent with consumer preferences. Among CQAs, only 5-CQA did not increase in concentration during the germination period, perhaps because it was being converted to 1,5-diCQA or used for lignin biosynthesis (Figure 2). Competing use of 5-CQA for lignin biosynthesis versus production of 1,5-diCQA during

sprouting could contribute to the less abundance of monoCQAs than diCQAs in sprouts. Regarding the distribution of CQAs in hypocotyls and cotyledons, higher amounts of CQAs in cotyledons (Figure 3C) were probably because the role of these compounds is in protecting against herbivores, pathogens, and harmful UV light to which cotyledons are exposed to in nature (Niggeweg et al., 2004; Cle et al., 2008; Leiss et al., 2009).

We performed a genome-wide analysis to increase our understanding of CQA biosynthesis and discovered three *HaHQTs* and two *HaHCTs*. These enzymes were well clustered with the previously characterized *HaHQTs* and *HaHCTs* from chicory and globe artichoke (Figure 4). These results demonstrated close evolutionary relationships and some level of domain consensus among different HQT and HCT isoforms between sunflower and other Asteraceae family plants (chicory and globe artichoke). Although predictions from our *in silico* analysis by ChloroP suggested *HaHQT3* would be localized to chloroplasts (Figure 4), we found that all *HaHQTs* fused with GFP localized in the cytoplasm (Figure 6); but we also observed fluorescence signals in the nucleus. This may result from protein diffusion

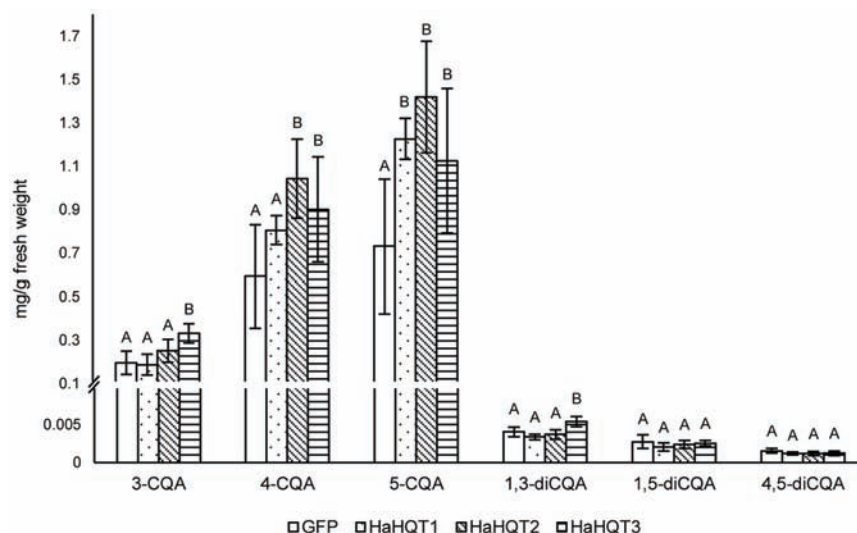


FIGURE 7 | HPLC profiling of caffeoylquinic acids in *Nicotiana benthamiana* leaves infiltrated with *HaHQTs*. Tissue extracts of *N. benthamiana* leaves infiltrated with GFP (control) and *HQTs* from sunflower (*HaHQT1*, *HaHQT2*, and *HaHQT3*) were analyzed by HPLC. Bars represent the mean values \pm standard deviation (SD) of five biologically independent replicates. Within each caffeoylquinic acid derivative, different alphabets indicate significant differences according to Duncan's multiple-range test ($p < 0.05$).

into the nucleus, consistent with the previous characterization of *HQT2* from globe artichoke (Moglia et al., 2016). A nucleocytoplasmic localization was in addition documented for other members of the BAHD superfamily, namely spermidine hydroxycinnamoyl transferases (Delporte et al., 2018).

Both *HQTs* and *HCTs* are involved in biosynthesis of CQAs in chicory (Legrand et al., 2016). Our expression analysis showed that *HaHQT2* was expressed at a dramatically higher level than other *HQTs* and *HCTs* in both hypocotyl and cotyledon tissues (Figure 5C). In addition, in cotyledons, the higher expression of *HaHQT2* coincided with higher content of CQAs than that in hypocotyls (Figures 5B,C). This suggests that, rather than the other genes, *HaHQT2* could be the main enzyme involved in CQA biosynthesis during sunflower sprout germination. Studies in tomato (Niggeweg et al., 2004), potato (Payyavula et al., 2015), globe artichoke (Moglia et al., 2016), and Japanese honeysuckle (*Lonicera japonica*; Zhang et al., 2017) reached a similar conclusion that *HQTs* have an important role in CQA biosynthesis. By contrast, based on functional characterization of *N. benthamiana*, Legrand et al. (2016) showed that *HCT1* from chicory played a major role in CQA biosynthesis. These results suggest that the mechanisms of CQA biosynthesis regulation could vary between closely related plant species.

To gain a better understanding of factors controlling the expression of *HaHQTs*, we searched the promoter regions of *HaHQT2* and *HaHQT3* for relevant regulatory elements. Multiple elements associated with stress response and hormonal signaling were found in the promoters of both genes. In addition, several MYB and Dof regulator elements were also found in the promoter regions of both genes. MYB transcription factor has been known to regulate CQA biosynthesis. Transient overexpression of MYB1 from eggplant (*Solanum melongena*)

in *N. benthamiana* could enhance the accumulation level of 5-CQA (Docimo et al., 2016). Moreover, 5-CQA content was significantly increased in leaves and fruit of transgenic tomato lines overexpressing Arabidopsis MYB12 (Pandey et al., 2015). The involvement of Dof transcription factor in phenylpropanoid biosynthesis has been reported. Arabidopsis Dof4.2 negatively affected flavonoid biosynthesis but positively regulated hydroxycinnamic acid biosynthesis (Skirycz et al., 2007). Since sunflower sprouts accumulate comparably high levels of CQAs, this sprout might be suitable for further identification of novel transcription factors controlling CQA biosynthesis. Nevertheless, the different expression levels of those two *HaHQTs* might be due to other regulating factors such as different positions/numbers of the cis-acting elements in the promoter regions. Further *in planta* functional analysis confirmed the role of all *HaHQTs* in CQA biosynthesis. Compared with control *N. benthamiana* plants infiltrated with GFP, leaves infiltrated with *HaHQTs* had higher 5-CQA content, and the highest increase (94%) occurred in leaves infiltrated by *HaHQT2* (Figure 7). These results suggest that *HaHQTs* are involved in the biosynthesis of 5-CQA, the most abundant monoCQA found in sunflower sprouts. Consistently, in *N. benthamiana* leaves infiltrated with artichoke *HQTs*, among other monoCQAs, 5-CQA content was mainly affected (Moglia et al., 2016). As for diCQAs, only 1,3-diCQA content was increased slightly by the infiltration of *N. benthamiana* leaves with *HaHQT3* construct. This observation was different from a study by Moglia et al. (2016) which found that *N. benthamiana* leaves infiltrated with artichoke *HQTs* had increased diCQA contents at much higher level than that of *HaHQT3*-transiently expressed *N. benthamiana* leaves. Because of this, we hypothesized that the increased amount of diCQA in *N. benthamiana* leaves infiltrated with artichoke *HQTs* might be due to the dramatically

increased level of monoCQAs (up to 500% for 5-CQA; Moglia et al., 2016), as diCQAs are synthesized from monoCQAs. However, in our study, we did not observe as high an amount of 3-CQA in the *HaHQT*s-infiltrated *N. benthamiana* leaves as Moglia et al. (2016) did, which might explain why we did not observe an increase in diCQA contents. In a study by Legrand et al. (2016), infiltration of *N. benthamiana* leaves with either *HQT1* or *HCT1* from chicory did not lead to a great increase in 3-CQA content (only up to ~19 and ~56%, respectively), which is similar to results from our study. Unfortunately, Legrand et al. (2016) did not mention any changes occurring in diCQA contents. Differences in the expression vectors used and in the growth stages and conditions of *N. benthamiana* between studies may partially explain why production of monoCQAs in *N. benthamiana* infiltrated with *HQT*s from different sources is variable. In addition, there may be differences in the catalytic efficiency of *HQT*s investigated in different studies, and this possibility should be investigated further.

Although sunflower sprouts accumulate much higher level of 1,5-diCQA than globe artichoke and chicory (Pandino et al., 2011; Willeman et al., 2014), our results did not provide a strong evidence to support the role of *HaHQT*s and *HaHCT*s in diCQA biosynthesis. Therefore, mechanisms behind biosynthesis of diCQAs remain unclear. It is possible that other acyltransferases might be involved in diCQA biosynthesis. Since the sunflower genome has recently been reported (Badouin et al., 2017), identification of additional acyltransferases in the genome, together with transcriptome analysis, might help in identifying novel candidate genes involved in diCQA production. So far, the only *HQT* enzyme known to produce both monoCQAs and diCQAs is from tomato, where it is localized to both the cytosol and vacuoles (Moglia et al., 2014). The amino acid residue His276 has been identified as partially responsible for the dual function of tomato *HQT* and mutation of this residue to Tyr decreases the production of diCQA *in vitro*. In chicory, globe artichoke, and sunflower, a Tyr residue is found at this position (**Supplementary Figure S2**). Therefore, it is unlikely that *HQT*s in these three species would have the dual function seen in tomato *HQT*.

In conclusion, we have reported biosynthesis of CQAs in sunflower sprouts for the first time. Sunflower sprouts are a rich source of monoCQAs and diCQAs, and *HaHQT2* was found to be the major isoform, which could be responsible for CQA biosynthesis during germination in both hypocotyls and cotyledons. Therefore, manipulation of this gene could positively affect CQA contents in sunflower sprouts. Thus, our results provide informative data, which could be applied to further biofortify sunflower sprouts as functional foods.

REFERENCES

- Badouin, H., Gouzy, J., Grassa, C. J., Murat, F., Staton, S. E., and Cottret, L. (2017). The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature* 546, 148–152. doi: 10.1038/nature22380
- Bannai, H., Tamada, Y., Maruyama, O., Nakai, K., and Miyano, S. (2002). Extensive feature detection of N-terminal protein sorting signals. *Bioinformatics* 18, 298–305. doi: 10.1093/bioinformatics/18.2.298
- Bashir, T., Mashwani, Z. R., Zahara, K., Haider, S., and Mudrikah, T. S. (2015). Chemistry, pharmacology and ethnomedicinal uses of *Helianthus annuus* (sunflower): a review. *Pure Appl. Biol.* 4, 226–235. doi: 10.19045/bspab.2015.42011
- Beekman, L., Tohver, T., Dardari, R., and Léguillette, R. (2011). Evaluation of suitable reference genes for gene expression studies in bronchoalveolar lavage cells from horses with inflammatory airway disease. *BMC Mol. Biol.* 12:5. doi: 10.1186/1471-2199-12-5

DATA AVAILABILITY

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

SS conceived the research. KC performed most of the experiments and analyzed the data. PB participated in genome-wide identification. KC, GK, PP, and SS interpreted the data and drafted the manuscript. All authors have read and approved the final manuscript.

FUNDING

This research work was financially supported by the 90th Anniversary of Chulalongkorn University (Ratchadaphisek Somphot Endowment Fund to KC) and Chulalongkorn research funding (GRU 6203023003-1 and CU-57-014-FW to SS).

ACKNOWLEDGMENTS

We would like to thank Tsuyoshi Nakagawa (Shimane University, Japan) for providing the Gateway® expression vector pGWB5, Sophien Kamoun (Sainsbury Laboratory, UK) for providing *A. tumefaciens* GV3101 carrying pJL3:p19, and George Lomonosoff (John Innes Centre, Norwich, UK) and Plant Bioscience Limited (Norwich, UK) for supplying the pEAQ vectors. We also thank Chadin Kulsing (Chulalongkorn University, Thailand) for assisting in LC–MS analysis and Kornlawat Tantivit for assisting in confocal microscopy. We appreciate a scholarship to develop research potential for the Department of Biochemistry, Faculty of Science, Chulalongkorn University, Ratchadaphisek Somphot Fund for Postdoctoral Fellowships, Chulalongkorn University (to GK and PP) and the Chulalongkorn Academic Advancement into its 2nd Century Project.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.00968/full#supplementary-material>

- Cartharius, K., Frech, K., Grote, K., Klocke, B., Haltmeier, M., Klingenhoff, A., et al. (2005). MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics* 21, 2933–2942. doi: 10.1093/bioinformatics/bti473
- Cle, C., Hill, L., Niggeweg, R., Martin, C., Guisez, Y., Prinsen, E., et al. (2008). Modulation of chlorogenic acid biosynthesis in *Solanum lycopersicum*; consequences for phenolic accumulation and UV-tolerance. *Phytochemistry* 69, 2149–2156. doi: 10.1016/j.phytochem.2008.04.024
- Comino, C., Hehn, A., Moglia, A., Menin, B., Bourgaud, F., Lanteri, S., et al. (2009). The isolation and mapping of a novel hydroxycinnamoyltransferase in the globe artichoke chlorogenic acid pathway. *BMC Plant Biol.* 9:30. doi: 10.1186/1471-2229-9-30
- Comino, C., Lanteri, S., Portis, E., Acquadro, A., Romani, A., Hehn, A., et al. (2007). Isolation and functional characterization of a cDNA coding a hydroxycinnamoyltransferase involved in phenylpropanoid biosynthesis in *Cynara cardunculus* L. *BMC Plant Biol.* 7:14. doi: 10.1186/1471-2229-7-14
- de Sotillo, D. V. R., and Hadley, M. (2002). Chlorogenic acid modifies plasma and liver concentrations of cholesterol, triacylglycerol, and minerals in (fa/fa) Zucker rats. *J. Nutr. Biochem.* 13, 717–726. doi: 10.1016/S0955-2863(02)00231-0
- Delparte, M., Bernard, G., Legrand, G., Hielscher, B., Lanoue, A., Molinié, R., et al. (2018). A BAHD neofunctionalization promotes tetrahydroxycinnamoyl spermine accumulation in the pollen coat of the Asteraceae family. *J. Exp. Bot.* 69, 5355–5371. doi: 10.1093/jxb/ery320
- Docimo, D., Francese, G., Ruggiero, A., Batelli, G., De Palma, M., Bassolino, L., et al. (2016). Phenylpropanoids accumulation in eggplant fruit: characterization of biosynthetic genes and regulation by a MYB transcription factor. *Front. Plant Sci.* 6:1233. doi: 10.3389/fpls.2015.01233
- Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* 300, 1005–1016. doi: 10.1006/jmbi.2000.3903
- Emanuelsson, O., Nielsen, H., and von Heijne, G. (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci.* 8, 978–984. doi: 10.1110/ps.8.5.978
- Escamilla-Treviño, L. L., Shen, H., Hernandez, T., Yin, Y., Xu, Y., and Dixon, R. A. (2014). Early lignin pathway enzymes and routes to chlorogenic acid in switchgrass (*Panicum virgatum* L.). *Plant Mol. Biol.* 84, 565–576. doi: 10.1007/s11103-013-0152-y
- Fowler, M. W. (2006). Plants, medicines and man. *J. Sci. Food Agric.* 86, 1797–1804. doi: 10.1002/jsfa.2598
- Gu, R., Dou, G., Wang, J., Dong, J., and Meng, Z. (2007). Simultaneous determination of 1,5-dicaffeoylquinic acid and its active metabolites in human plasma by liquid chromatography-tandem mass spectrometry for pharmacokinetic studies. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* 852, 85–91. doi: 10.1016/j.jchromb.2006.12.055
- Guo, S., Ge, Y., and Na Jom, K. (2017). A review of phytochemistry, metabolite changes, and medicinal uses of the common sunflower seed and sprouts (*Helianthus annuus* L.). *Chem. Cent. J.* 11:95. doi: 10.1186/s13065-017-0328-7
- Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41, 95–98.
- Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., et al. (2007). WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* 35, 585–587. doi: 10.1093/nar/gkm259
- Kojima, M., and Kondo, T. (1985). An enzyme in sweet potato root which catalyzes the conversion of chlorogenic acid, 3-caffeoylquinic acid, to isochlorogenic acid, 3,5-dicaffeoylquinic acid. *Agric. Biol. Chem.* 49, 2467–2469.
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054
- Lallemant, L., Zubietta, C., Lee, S., Wang, Y., Acajaoui, S., Timmins, J., et al. (2012). A structural basis for the biosynthesis of the major chlorogenic acids found in coffee. *Plant Physiol.* 160, 249–260. doi: 10.1104/pp.112.202051
- Lawal, U., Leong, S. W., Shaari, K., Ismail, I. S., Khatib, A., and Abas, F. (2016). α -Glucosidase inhibitory and antioxidant activities of different *Ipomoea aquatica* cultivars and LC-MS/MS profiling of the active cultivar. *J. Food Biochem.* 41:12303. doi: 10.1111/jfbc.12303
- Legrand, G., Delparte, M., Khelifi, C., Harant, A., Vuylsteker, C., Morchen, M., et al. (2016). Identification and characterization of five BAHD acyltransferases involved in hydroxycinnamoyl ester metabolism in chicory. *Front. Plant Sci.* 7:741. doi: 10.3389/fpls.2016.00741
- Leiss, K., Maltese, F., Choi, Y., Verpoorte, R., and Klinkhamer, P. (2009). Identification of chlorogenic acid as a resistance factor for thrips in *Chrysanthemum*. *Plant Physiol.* 150, 1567–1575. doi: 10.1104/pp.109.138131
- Lindbo, J. A. (2007). High-efficiency protein expression in plants from agroinfection-compatible tobacco mosaic virus expression vectors. *BMC Biotechnol.* 7:52. doi: 10.1186/1472-6750-7-52
- Maresh, V., Million-Rousseau, R., Ullmann, P., Chabrilange, N., Bustamante, J., Mondolot, L., et al. (2007). Functional characterization of two p-coumaroyl ester 3'-hydroxylase genes from coffee tree: evidence of a candidate for chlorogenic acid biosynthesis. *Plant Mol. Biol.* 64, 145–159. doi: 10.1007/s11103-007-9141-3
- McDougall, B., King, P. J., Wu, B. W., Hostomsky, Z., Reinecke, M. G., and Robinson, W. E. Jr. (1998). Dicafeoylquinic and dicafeoyltartaric acids are selective inhibitors of human immunodeficiency virus type 1 integrase. *Antimicrob. Agents Chemother.* 42, 140–146. doi: 10.1128/AAC.42.1.140
- McWilliam, H., Li, W., Uludag, M., Squizzato, S., Mi, Y., Buso, N., et al. (2013). Analysis tool web services from the EMBL-EBI. *Nucleic Acids Res.* 41, 597–600. doi: 10.1093/nar/gkt376
- Menin, B., Comino, C., Moglia, A., Dolzhenko, Y., Portis, E., and Lanteri, S. (2010). Identification and mapping of genes related to caffeoylquinic acid synthesis in *Cynara cardunculus* L. *Plant Sci.* 179, 338–347. doi: 10.1016/j.plantsci.2010.06.010
- Moglia, A., Acquadro, A., Eljounaidi, K., Milani, A. M., Cagliero, C., Rubiolo, P., et al. (2016). Genome-wide identification of BAHD acyltransferases and in vivo characterization of HQT-like enzymes involved in caffeoylquinic acid synthesis in globe artichoke. *Front. Plant Sci.* 7:1424. doi: 10.3389/fpls.2016.01424
- Moglia, A., Comino, C., Portis, E., Acquadro, A., De Vos, R., Beekwilder, J., et al. (2009). Isolation and mapping of a C3'H gene (CYP98A49) from globe artichoke, and its expression upon UV-C stress. *Plant Cell Rep.* 28, 963–974. doi: 10.1007/s00299-009-0695-1
- Moglia, A., Lanteri, S., Comino, C., Hill, L., Knevit, D., Cagliero, C., et al. (2014). Dual catalytic activity of hydroxycinnamoyl-coenzyme A quinate transferase from tomato allows it to moonlight in the synthesis of both mono- and dicafeoylquinic acids. *Plant Physiol.* 166, 1777–1787. doi: 10.1104/pp.114.251371
- Nakagawa, T., Kurose, T., Hino, T., Tanaka, K., Kawamukai, M., Niwa, Y., et al. (2007). Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *J. Biosci. Bioeng.* 104, 34–41. doi: 10.1263/jbb.104.34
- Naveed, M., Hejazi, V., Abbas, M., Kamboh, A. A., Khan, G. J., Shumzaid, M., et al. (2018). Chlorogenic acid (CGA): a pharmacological review and call for further research. *Biomed. Pharmacother.* 97, 67–74. doi: 10.1016/j.bioph.2017.10.064
- Niggeweg, R., Michael, A., and Martin, C. (2004). Engineering plants with increased levels of the antioxidant chlorogenic acid. *Nat. Biotechnol.* 22, 746–754. doi: 10.1038/nbt966
- Ochogavia, A. C., Novello, M. A., Picardi, L. A., and Nestares, G. M. (2017). Identification of suitable reference genes by quantitative real-time PCR for gene expression normalization in sunflower. *Plant Omics* 10, 210–218. doi: 10.21475/poj.10.04.17.pne831
- Pajak, P., Socha, R., Galkowska, D., Rożnowski, J., and Fortuna, T. (2014). Phenolic profile and antioxidant activity in selected seeds and sprouts. *Food Chem.* 143, 300–306. doi: 10.1016/j.foodchem.2013.07.064
- Pandey, A., Misra, P., Choudhary, D., Yadav, R., Goel, R., Bhambhani, S., et al. (2015). AtMYB12 expression in tomato leads to large scale differential modulation in transcriptome and flavonoid content in leaf and fruit tissues. *Sci. Rep.* 5:12412. doi: 10.1038/srep12412
- Pandino, G., Lombardo, S., and Mauromicale, G. (2011). Mineral profile in globe artichoke as affected by genotype, head part and environment. *J. Sci. Food Agric.* 91, 302–308. doi: 10.1002/jsfa.4185
- Payyavula, R. S., Shakya, R., Sengoda, V. G., Munyaneza, J. E., Swamy, P., and Navarre, D. A. (2015). Synthesis and regulation of chlorogenic acid in potato: rerouting phenylpropanoid flux in HQT-silenced lines. *Plant Biotechnol. J.* 13, 551–564.
- Peyret, H., and Lomonosoff, G. P. (2013). The pEAQ vector series: the easy and quick way to produce recombinant proteins in plants. *Plant Mol. Biol.* 83, 51–58. doi: 10.1007/s11103-013-0036-1
- Salomone, F., Galvano, F., and Volti, G. L. (2017). Molecular bases underlying the hepatoprotective effects of coffee. *Nutrients* 9:85. doi: 10.3390/nu9010085

- Schmittgen, T. D., and Livak, K. J. (2008). Analyzing real-time PCR data by the comparative C_T method. *Nat. Protoc.* 3, 1101–1108. doi: 10.1038/nprot.2008.73
- Skirycz, A., Jozefczuk, S., Stobiecki, M., Muth, D., Zanor, M. I., Witt, I., et al. (2007). Transcription factor AtDOF4;2 affects phenylpropanoid metabolism in *Arabidopsis thaliana*. *New Phytol.* 175, 425–438. doi: 10.1111/j.1469-8137.2007.02129.x
- Sperschneider, J., Catanzariti, A. M., DeBoer, K., Petre, B., Gardiner, D. M., Singh, K. B., et al. (2017). LOCALIZER: Subcellular localization prediction of both plant and effector proteins in the plant cell. *Sci. Rep.* 7:44598. doi: 10.1038/srep44598
- Sun, Z., Chen, J., Ma, J., Jiang, Y., Wang, M., Ren, G., et al. (2012). Cynarin-rich sunflower (*Helianthus annuus*) sprouts possess both antiglycative and antioxidant activities. *J. Agric. Food Chem.* 60, 3260–3265. doi: 10.1021/jf300737y
- Taylor, S. C., Laperriere, G., and Germain, H. (2017). Droplet Digital PCR versus qPCR for gene expression analysis with low abundant targets: from variable nonsense to publication quality data. *Sci. Rep.* 7:2409. doi: 10.1038/s41598-017-02217-x
- Tuominen, L. K., Johnson, V. E., and Tsai, C.-J. (2011). Differential phylogenetic expansions in BAHD acyltransferases across five angiosperm taxa and evidence of divergent expression among *Populus* paralogues. *BMC Genomics* 12:236. doi: 10.1186/1471-2164-12-236
- Weisz, G. M., Kammerer, D. R., and Carle, R. (2009). Identification and quantification of phenolic compounds from sunflower (*Helianthus annuus* L.) kernels and shells by HPLC-DAD/ESI-MSn. *Food Chem.* 115, 758–765. doi: 10.1016/j.foodchem.2008.12.074
- Willeman, H., Hance, P., Fertin, A., Voedts, N., Duhal, N., Goossens, J. F., et al. (2014). A method for the simultaneous determination of chlorogenic acid and sesquiterpene lactone content in industrial chicory root foodstuffs. *Sci. World J.* 2014, 1–11. doi: 10.1155/2014/583180
- Xu, J., Hu, Q., and Liu, Y. (2012). Antioxidant and DNA-protective activities of chlorogenic acid isomers. *J. Agric. Food Chem.* 60, 11625–11630. doi: 10.1021/jf303771s
- Yu, X.-H. H., Gou, J.-Y. Y., and Liu, C.-J. J. (2009). BAHD superfamily of acyl-CoA dependent acyltransferases in *Populus* and *Arabidopsis*: bioinformatics and gene expression. *Plant Mol. Biol.* 70, 421–442. doi: 10.1007/s11103-009-9482-1
- Zhang, J., Wu, M., Li, W., and Bai, G. (2017). Regulation of chlorogenic acid biosynthesis by hydroxycinnamoyl CoA quinate hydroxycinnamoyl transferase in *Lonicera japonica*. *Plant Physiol. Biochem.* 121, 74–79.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Cheevarungnapakul, Khaksar, Panpetch, Boonjing and Sirikantaramas. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Origin and Evolution of Plant Flavonoid Metabolism

Keiko Yonekura-Sakakibara*, Yasuhiro Higashi and Ryo Nakabayashi

RIKEN Center for Sustainable Resource Science, Yokohama, Japan

During their evolution, plants have acquired the ability to produce a huge variety of compounds. Unlike the specialized metabolites that accumulate in limited numbers of species, flavonoids are widely distributed in the plant kingdom. Therefore, a detailed analysis of flavonoid metabolism in genomics and metabolomics is an ideal way to investigate how plants have developed their unique metabolic pathways during the process of evolution. More comprehensive and precise metabolite profiling integrated with genomic information are helpful to emerge unexpected gene functions and/or pathways. The distribution of flavonoids and their biosynthetic genes in the plant kingdom suggests that flavonoid biosynthetic pathways evolved through a series of steps. The enzymes that form the flavonoid scaffold structures probably first appeared by recruitment of enzymes from primary metabolic pathways, and later, enzymes that belong to superfamilies such as 2-oxoglutarate-dependent dioxygenase, cytochrome P450, and short-chain dehydrogenase/reductase modified and varied the structures. It is widely accepted that the first two enzymes in flavonoid biosynthesis, chalcone synthase, and chalcone isomerase, were derived from common ancestors with enzymes in lipid metabolism. Later enzymes acquired their function by gene duplication and the subsequent acquisition of new functions. In this review, we describe the recent progress in metabolomics technologies for flavonoids and the evolution of flavonoid skeleton biosynthetic enzymes to understand the complicate evolutionary traits of flavonoid metabolism in plant kingdom.

Keywords: secondary metabolites, flavonoid, polyketide synthase, 2-oxoglutarate-dependent dioxygenase, cytochrome P450, short-chain dehydrogenase/reductase, plant

OPEN ACCESS

Edited by:

Kevin Davies,
The New Zealand Institute for Plant
and Food Research Ltd., New Zealand

Reviewed by:

Stefan Martens,
Fondazione Edmund Mach, Italy
John A. Morgan,
Purdue University, United States

*Correspondence:

Keiko Yonekura-Sakakibara
yskeiko@riken.jp

Specialty section:

This article was submitted to
Plant Metabolism and Chemodiversity,
a section of the journal
Frontiers in Plant Science

Received: 26 April 2019

Accepted: 08 July 2019

Published: 02 August 2019

Citation:

Yonekura-Sakakibara K, Higashi Y and
Nakabayashi R (2019) The Origin and
Evolution of Plant Flavonoid
Metabolism. *Front. Plant Sci.* 10:943.
doi: 10.3389/fpls.2019.00943

INTRODUCTION

Plants have the ability to produce a huge variety of metabolites. Over 1,000,000 metabolites are predicted to be present in the entire plant kingdom (Afendi et al., 2012). Most of these are secondary metabolites (also referred to as specialized metabolites) that play a wide range of physiological and ecological roles including defense against herbivores and pathogens, attractants for pollinators and seed carriers, and signaling. During the long process of evolution, plants have gained, expanded, and sometimes lost their capabilities to produce this huge array of metabolites, which provides the adaptive mechanisms needed for survival in changing environments.

Flavonoids form one of the major groups of specialized metabolites, and include over 9,000 compounds (Williams and Grayer, 2004; Anderson and Markham, 2006). According to Nomenclature of flavonoids (IUPAC Recommendations, 2017), the term “flavonoid” is applied to (1) compounds structurally based on derivatives of a phenyl-substituted propylbenzene having a C15 skeleton, (2) compounds with a C16 skeleton that are phenyl-substituted propylbenzene derivatives (rotenoids), (3) flavonolignans based on derivatives of phenyl-substituted propylbenzene condensed with C₆-C₃ lignan precursors (Rauter et al., 2018). In a

restricted sense, the term “flavonoid” is used only for those compounds with a C₆-C₃-C₆ carbon framework exhibiting the structure of a chromane or that of a chromene such as flavans, flavones, flavonols, and anthocyanidins. Chalcones, dihydrochalcones, and aurones are flavonoids in a broad sense, but not in a limited sense.

Flavonoids, including chalcones, flavones, flavonols, anthocyanins, and proanthocyanidins, are widely distributed in the plant kingdom, and their metabolic pathways have been extensively studied using both biochemical and molecular biological techniques. Until recently, it was believed that liverworts and mosses were the oldest flavonoid-producing plants (Rausher, 2006; Bowman et al., 2017). The genes encoding enzymes in the phenylpropanoid biosynthetic pathway, including the first two enzymes for flavonoid biosynthesis (chalcone synthase and chalcone isomerase) had not been found in the algal genera *Chlamydomonas*, *Micromonas*, *Ostreococcus*, and *Klebsormidium*, although genes encoding enzymes in the shikimate pathway were found in algae, liverworts, mosses, lycophytes, ferns and horsetails, gymnosperms, and angiosperms (Bowman et al., 2017). However, flavones, isoflavones, and flavonols were detected in microalgae from five different evolutionary lineages (*Cyanobacteria*, *Rhodophyta*, *Chlorophyta*, *Haptophyta*, and *Ochrophyta*) using ultra-high performance liquid chromatography with tandem mass spectrometry (Goiris et al., 2014). This suggests that plants may have acquired the ability to produce flavonoids earlier than we previously thought. Furthermore, in extant plants flavonoids play important roles as ultraviolet-B (UV-B) protectants, pigments that attract pollinators, phytoalexins, signaling molecules, and regulators of auxin transport and fertility (Gould and Lister, 2006). It has been proposed that defense against UV irradiation and regulation of plant hormone action were the original functions of flavonoids in the earliest flavonoid producing plants (Stafford, 1991; Shirley, 1996; Rausher, 2006). These functions have been considerably diversified during plant evolution. Thus, the study of flavonoids is a useful approach to understand how plants acquired the ability to produce specialized metabolites, and then build the metabolic pathways needed to produce such a huge variety of metabolites during the course of their evolution. This study will shed light on the relationships between genes/proteins and metabolites, and between the metabolites and their physiological functions. In this review, we describe the structural diversity of flavonoids distributed in the plant kingdom, and how plants acquired flavonoid biosynthetic genes during their evolution.

METABOLOMICS SHEDS LIGHT ON THE EVOLUTION OF FLAVONOIDS

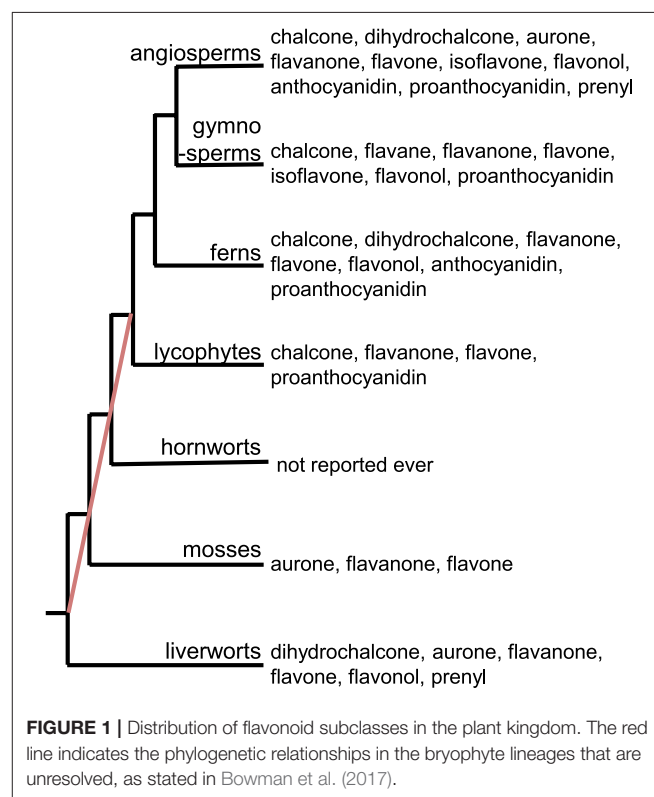
Distribution of Flavonoids in the Plant Kingdom

It is estimated that there are over 9,000 flavonoids in the plant kingdom (Williams and Grayer, 2004; Anderson and Markham, 2006). Research on flavonoids has shown that they are distributed across the plant kingdom, including angiosperms, gymnosperms, and pteridophytes (Harborne, 1988; Tohge et al., 2013). The abundance of information about flavonoids

in different species allows us to identify which flavonoid subclasses (e.g., chalcones, flavones, flavonols, anthocyanins, and proanthocyanidins) are found in each subgroup of plants (Figure 1). Flavone and flavanone are found in all plant groups except for hornworts. To our knowledge, no flavonoids have been reported in hornworts. As plant groups have evolved and diversified, so have the flavonoid subclasses produced within each group. For example, the flavonoid aglycones are most diversified in the angiosperms. In addition to flavanones and flavones, chalcones, flavonols, and proanthocyanidins are found in multiple groups. Interestingly, the prenylflavonoids are found in both liverworts and angiosperms: while more than 1,000 prenylflavonoids have been isolated from legumes (Yazaki et al., 2009), prenyldihydrochalcone has been found in the liverworts *Radula variabilis* and *Radula* spp. (Asakawa et al., 1978, 1982). These data suggest that the two groups of plants gained the ability to produce prenylflavonoids independently, or that many groups have lost the ability to produce prenylflavonoids during their evolution. Flavonoid molecules provide concrete evidence for the existence of the corresponding flavonoid biosynthetic genes in the plants. Analytical approaches for identifying flavonoids is therefore important to understand the evolution of flavonoid metabolism in the plant kingdom.

Cutting-Edge Metabolomics Technologies Contribute to Understanding the Evolution of Flavonoid Metabolism

So far, chromatographic and spectroscopic approaches have been used to analyze the structures of flavonoids and to reveal their chemical diversity. Previously, paper chromatography,



thin layer chromatography, column chromatography, and liquid chromatography (LC) were the main techniques used to study flavonoids. Crude extracts containing flavonoids are obtained through liquid-liquid partitioning. For instance, flavonoid aglycones and their mono-glycosides can be mainly extracted in ethylacetate, whereas the flavonoid di- or tri-glycosides are extracted in *n*-butanol. Chromatographic techniques are then used to purify the flavonoids from the ethylacetate or *n*-butanol fractions, in several steps. Finally, the isolated flavonoids are analyzed by nuclear magnetic resonance and mass spectrometry to elucidate their structures. This is an unequivocal and straightforward approach to identifying flavonoids, however, it is time-consuming.

With the development of metabolomics technologies, flavonoids can now be analyzed much more accurately and precisely than before. LC-tandem mass spectrometry (LC-MS/MS) has become the preferred approach for analyzing flavonoids. It had been considered that there was no flavonoids in algae. However, identification of flavonoids such as intermediates and end products by an LC-MS/MS approach proved the existence of the flavonoid biosynthetic ability in microalgae (Goiris et al., 2014). It suggests that cutting-edge metabolomics technologies in every single plant species can unveil the flavonoids that have been missed.

The integration of LC-MS/MS with cheminformatics approaches provides a powerful tool for surveying flavonoid diversity in a high-throughput way (Tsugawa et al., 2016; Akimoto et al., 2017). Glycosylated, acylated, and prenylated flavonoid molecules and their aglycones can be separated using simple combinations of solvents and LC columns. The separated molecules are then ionized for MS/MS analysis. The product ions are derived from their precursors through the cleavage of ether and ester bonds or even the prenyl moiety. These simple fragmentation steps allow us to estimate the structures of the flavonoids. Cheminformatic tools are used for high-throughput and (semi-)automatic data analysis. The results are generally classified into four classes (levels) according to the guidelines of the Metabolomics Standards Initiative: Level 1, identification using authentic standard compounds; Level 2, annotation using public databases; Level 3, characterization by deciphering MS/MS data; and Level 4, unknown (not noise) (Sumner et al., 2007). The use of these levels reduces the production of false positive data.

An integrated approach using LC-MS/MS with ^{13}C (carbon) labeling and cheminformatics can be used to assign structures to metabolites whose structures have never been recognized before. In most of the general approaches, unknown metabolites remain uncharacterized. The integrative approach is used to identify the elemental compositions of the unknown metabolites based on the numbers of C atoms, determined by comparisons between ^{13}C - and non-labeled MS spectra. The MS/MS spectra of the unknown metabolites are then subjected to fragment set enrichment analysis (FSEA) to determine candidates for each metabolite class. This integrative approach was used to characterized 1,133 metabolites, including flavonoids, in 12 angiosperm species (Tsugawa et al., 2019). The results for the characterized flavonoids are summarized in **Table 1**, along with the flavonoids found in the

liverwort *Marchantia polymorpha*, which were analyzed using the same method (Kubo et al., 2018). Our analyses using LC-MS/MS with ^{13}C labeling and cheminformatics led to the identification of more flavonoid molecules especially in the group of flavonol and flavone. Prenylated flavonoids have been analyzed by LC-MS/MS less extensively than other flavonoid groups, although a chemical assignment strategy is effective for profiling prenylated flavonoids. It is probably due to the lack of authentic standards and publicly available MS/MS spectra. The exact mass of a known structure or one that is presumed to be prenylated can also be useful for profiling prenylated flavonoids in cases where the ontologies for prenylated flavonoids does not work. This means that the observed MS/MS spectra may not be matched to the library derived from the FSEA.

Similarly, recent studies suggest that the tricin metabolism seems to specifically evolve in monocot plants. Tricin serves as a lignin co-monomer and cross couples with monolignols and γ -*p*-coumaroylated monolignols upon cell wall lignification. Recently, tricin was identified as a nucleation site for lignification in grasses (Lan et al., 2015, 2016a; Lam et al., 2017). Poales plants including *Avena sativa*, *Brachipodium distachyon*, *Oryza sativa*, *Triticum durum*, and *Zea mays* accumulate tricin, as does the Fabales plant *Medicago sativa* (Lan et al., 2016b). An integrative approach was used to detect derivatives of apigenin, luteolin, and tricin in *O. sativa*, *Z. mays*, and *M. truncatula*, and it appears that the total tricin content is much higher than the extractable tricin content in these plants. The integrative approach with an FSEA may be useful for further understanding the role of tricin and its derivatives in the cell wall.

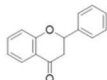
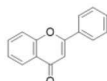
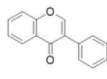
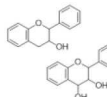
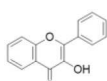
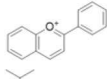
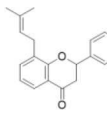
LC-MS/MS-based metabolome analyses have been used to comprehensively analyze the flavonoid molecules in individual plant species. Recent technological developments allow us to perform high-throughput LC-MS/MS analyses. By tracing the ions derived from the aglycones and modified parts of the molecules, or by finding MS/MS similarities, the chemical assignment of flavonoids is easily performed in (semi-)automatic ways with cheminformatic tools across plant species.

Biosynthetic genes for the formation of aglycone or modification can be predicted on the basis of the structure of flavonoids. Functional genomics using both cutting-edge sequencer and more comprehensive and accurate flavonoid profiling allow us to address the evolutionary questions of when and why flavonoids appeared and the biosynthetic pathway are diversified in plants. Furthermore, along with the progress of considerable plant genome projects, it will provide valuable clues to understand the evolutionary traits of flavonoid metabolism in plant kingdom as inconsistencies in the relationships between metabolites and genes that we mentioned in later sections.

THE FLAVONOID BIOSYNTHETIC PATHWAYS

The first step in flavonoid biosynthesis is catalyzed by chalcone synthase (CHS) (**Figure 2**). The substrates *p*-coumaroyl-CoA, derived from the cinnamate/monolignol (phenylpropanoid) pathway, and malonyl-CoA from the acetate-malonate

TABLE 1 | Flavonoids experimentally characterized by LC-MS/MS in 12 angiosperm species and *Marchantia polymorpha*.

Group	Representative structure		Name															
			Marchantiales			Asparagales		Poales		Solanales		Gentianales		Brassicales		Fabales		
			Marchantia polymorpha	Allium cepa	Oryza Sativa	Zea mays	Lycopersicon esculentum	Nicotiana tabacum	Solanum tuberosum	Optiorrhiza pumila	Arabidopsis thaliana	Glycine max	Glycyrrhiza glabra	Glycyrrhiza uralensis	Medicago truncatula			
Flavanone		Eriodictyol, hesperetin, isookanin, liquiritigenin, naringenin, sakuranetin, etc.	0 (0)	1 (0)	0 (2)	0 (0)	2 (0)	0 (0)	0 (0)	0 (0)	0 (0)	2 (0)	10 (5)	7 (17)	1 (0)			
Flavone		Apigenin, baicalein, chrysoeriol, luteolin, scutellarein, tricetin, etc.	5 (3)	1 (5)	30 (19)	23 (17)	2 (0)	2 (13)	3 (0)	1 (0)	0 (0)	10 (4)	15 (2)	11 (5)	12 (0)			
Isoflavone		Afromosin, biochanin, daizein, formononetin, genistein, etc.	0 (0)	0 (0)	0 (0)	1 (0)	0 (0)	0 (6)	0 (0)	8 (0)	0 (0)	18 (18)	14 (8)	17 (3)	5 (0)			
Flavanol/flavandiol		Catechin, procyanidin	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	2 (0)	2 (0)	0 (0)			
Flavonol		Gossypetin, isorhamnetin, kaempferide, kaempferol, laricitrin, myricetin, patuletin, quercetin, syringetin	0 (0)	10 (3)	2 (0)	6 (3)	7 (0)	3 (1)	10 (3)	0 (0)	9 (1)	6 (9)	19 (2)	19 (0)	15 (0)			
Anthocyanidin		Cyanidin, delphinidin, petunidin, riccionidin, etc.	1 (2)	0 (10)	0 (0)	1 (0)	1 (0)	0 (0)	0 (6)	0 (0)	0 (3)	1 (0)	2 (0)	2 (0)	0 (0)			
Prenylated flavonoid		Bolusanthol, gancaonin, glabrol, kievitone, licoisoflavone, sigmoidin, etc.	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (4)	13 (49)	10 (40)	0 (0)			

The numbers of flavonoids characterized by Tsugawa et al. (2019) are shown. The numbers of flavonoids stored in the Dictionary of Natural Products are shown in parentheses. A conventional type of reverse phase LC-MS/MS was used for these analyses. On the basis of the ontology in the FSEA, the flavonoids are summarized into seven groups. Flavonoids with the prenyl moiety are grouped together as the prenylated flavonoid group. Prenylations (29 patterns) occur in chalcones/dihydrochalcones, flavanones, flavones, flavonols/dihydroflavonols, and isoflavones (Barron and Ibrahim, 1996). Flavonoids consisting of unmodified aglycones and glycosylated, acylated, etherated, and/or prenylated flavonoids were characterized in this study.

(polyketide) pathway are converted by CHS to naringenin chalcone. The stereospecific cyclization of naringenin chalcone to naringenin is catalyzed by chalcone isomerase (CHI). This step can also proceed spontaneously.

Naringenin is a general precursor for flavonols, anthocyanins, proanthocyanidins, flavones, and isoflavones. It is converted to dihydrokaempferol by flavanone 3-hydroxylase (F3H) (also referred to as flavanone 3 β -hydroxylase). Flavonoid 3'-hydroxylase (F3'H) and flavonoid F3'5'H-hydroxylase (F3'5'H) catalyze the hydroxylation of the C3' and C3'/C5' positions of dihydrokaempferol, respectively. Dihydroflavonol 4-reductase (DFR) catalyzes the reduction of the dihydroflavonols to leucoanthocyanidins, which are further converted to anthocyanidins by leucoanthocyanidin dioxygenase/anthocyanidin synthase (LDOX/ANS). The leucoanthocyanidins and anthocyanidins are reduced to flavan 3-ols (e.g., catechin and epicatechin) by leucoanthocyanidin reductase (LAR) and anthocyanidin reductase (ANR), respectively. The dihydroflavonols are also converted to flavonols by flavonol synthase (FLS).

Flavones are synthesized from naringenin by flavone synthase I (FNS I) or flavone synthase II (FNS II) (Martens and Mithofer, 2005). Flavanone 2-hydroxylase (F2H) catalyzes the hydroxylation of flavanones (including naringenin) to 2-hydroxyflavanones, which are subsequently converted to flavones, possibly by an unknown dehydratase (Akashi et al., 1998).

Isoflavone synthase (2-hydroxyisoflavanone synthase, IFS) catalyzes the first step of isoflavone biosynthesis. IFS converts flavanones (e.g., liquiritigenin and naringenin) to 2-hydroxyisoflavanones, then the 2-hydroxyisoflavanones are dehydrated to isoflavones by 2-hydroxyisoflavanone dehydratase (HID) (Akashi et al., 2005).

The flavonoid skeletons mentioned above are highly modified by enzymes such as glycosyltransferases (GTs), acyltransferases, and methyltransferases. Here, we focus on the enzymes involved in the biosynthesis of the flavonoid skeleton molecules.

EVOLUTIONARY HISTORY OF THE FLAVONOID BIOSYNTHETIC PATHWAYS

Generally, the enzymes in secondary metabolic pathways have been derived from those involved in primary metabolism (Pichersky and Gang, 2000; Moghe and Last, 2015; Carrington et al., 2018), and the enzymes in flavonoid metabolism are no exception. CHS and CHI are derived from enzymes involved in fatty acid metabolism: β -ketoacyl ACP synthase and fatty acid binding protein, respectively (Ngaki et al., 2012; Weng and Noel, 2012a). CHS and β -ketoacyl ACP synthase are members of the type III polyketide synthase (PKS) family (Schuz et al., 1983). The enzymes in later steps of the flavonoid biosynthetic pathways belong to families such as the 2OGD, CYP, and short-chain dehydrogenase/reductase (SDR) superfamilies (Table 2). The members of these superfamilies are widely involved in primary and secondary metabolisms, suggesting that plants have acquired the enzyme functions in later biosynthetic pathways

TABLE 2 | The origin or gene family of flavonoid biosynthetic genes.

Gene	Abbreviation	Origin/gene family
<i>Chalcone synthase</i>	CHS	Type III polyketide synthase
<i>Chalcone isomerase</i>	CHI	Fatty acid binding protein
<i>Chalcone isomerase-like</i>	CHIL	Fatty acid binding protein
<i>Flavanone 3-hydroxylase</i>	F3H	2-oxoglutarate-dependent dioxygenase (DOXC28)
<i>Flavone synthase I</i>	FNS I	2-oxoglutarate-dependent dioxygenase (DOXC28)
<i>Flavone synthase II</i>	FNS II	Cytochrome P450 (CYP93B)
<i>Isoflavone synthase</i>	IFS	Cytochrome P450 (CYP93C)
<i>Flavonoid 3'-hydroxylase</i>	F3'H	Cytochrome P450 (CYP75)
<i>Flavonoid 3',5'-hydroxylase</i>	F3'5'H	Cytochrome P450 (CYP75)
<i>Flavonol synthase</i>	FLS	2-oxoglutarate-dependent dioxygenase (DOXC47)
<i>Dihydroflavonol 4-reductase</i>	DFR	Short-chain dehydrogenase (SDR108E)
<i>Leucoanthocyanidin dioxygenase/anthocyanidin synthase</i>	LDOX/ANS	2-oxoglutarate-dependent dioxygenase (DOXC47)
<i>Anthocyanidin reductase</i>	ANR	Short-chain dehydrogenase (SDR108E)
<i>Leucoanthocyanidin reductase</i>	LAR	Short-chain dehydrogenase (SDR460A)
<i>Glycosyltransferase</i>	GT	Family 1 glycosyltransferase, glycoside hydrolase family 1
<i>Acyltransferase</i>	AT	BAHD acyltransferase, serine carboxypeptidase-like acyltransferase
<i>Methyltransferase</i>	MT	Methyltransferase

via gene duplication and evolution of new functions for the duplicated gene products.

Based on the distribution patterns of the flavonoid subclasses, it has been suggested that the flavonoid biosynthetic pathways may have evolved via a series of steps, and that the first flavonoid biosynthetic enzymes were CHS, CHI, and F3H (Stafford, 1991; Rausher, 2006). It has also been proposed that CHS evolved first, followed by F3H and then CHI, because CHS catalyzes the first committed step in the pathways, and the step catalyzed by CHI can also proceed spontaneously (Rausher, 2006). Flavonoids are widely distributed among mosses, liverworts, and vascular plants, but are not found in hornworts. Algae generally contain no flavonoids (Rausher, 2006), but they have been found in a few evolutionary divergent lineages of microalgae (Goiris et al., 2014). These observations suggest that the ability to produce flavonoids may have evolved multiple times, or that the ability was widely lost during evolutionary processes. An analysis of the evolutionary rates of six genes involved in anthocyanin biosynthesis indicated that the upstream genes (*CHS*, *CHI*, and *F3H*) evolved more slowly than the downstream genes (*DFR*, *LDOX/ANS*, and *UDP-glucose:flavonoid 3-O-glucosyltransferase*) (Rausher et al., 1999). The upstream genes may be evolutionarily constrained due to their profound effects on the pathways. In

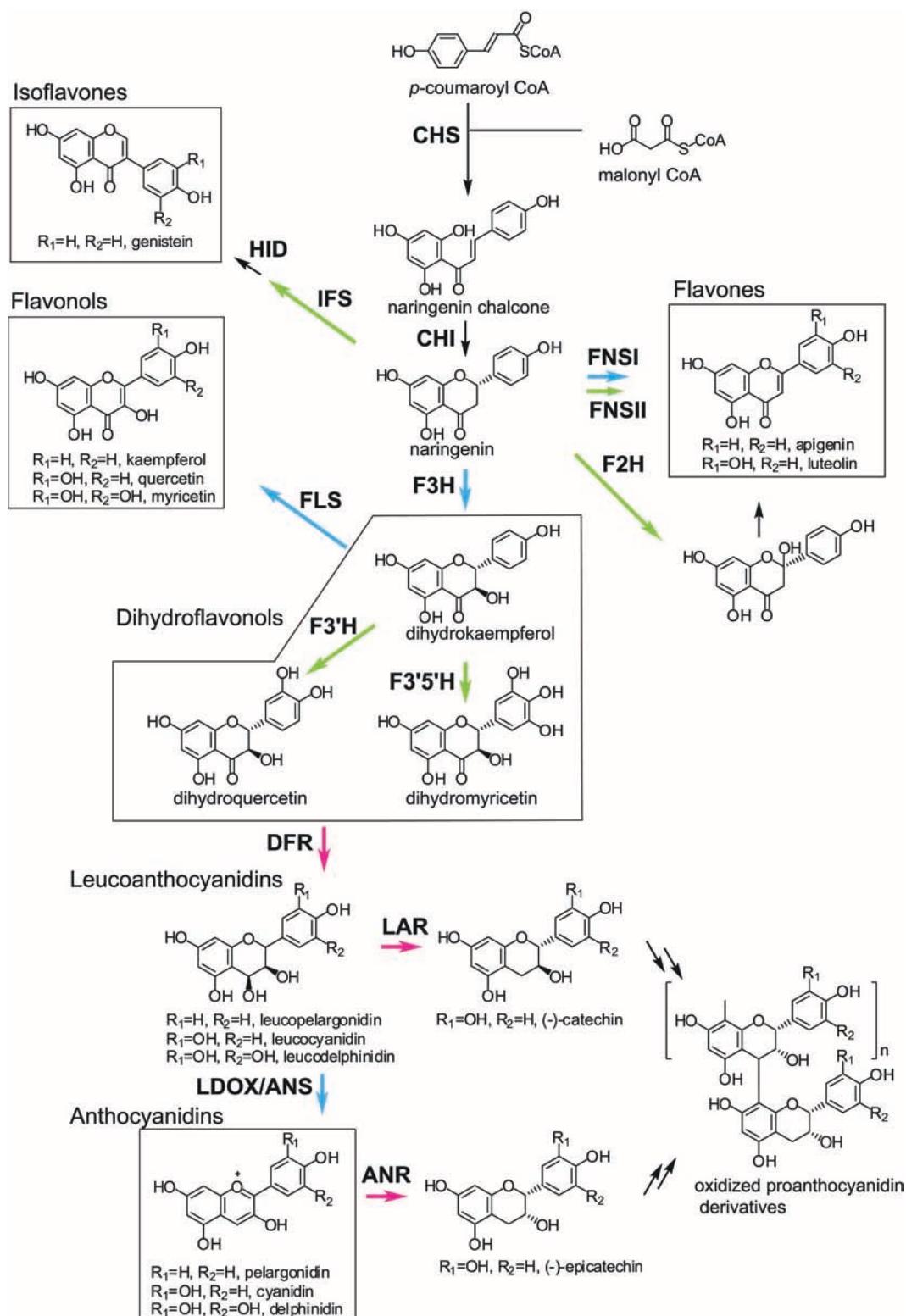


FIGURE 2 | General flavonoid biosynthetic pathways in plants. The arrows in green, blue, and magenta indicate enzymes in the CYP, 2OGD, and SDR superfamilies, respectively. ANR, anthocyanidin reductase; ANS, anthocyanidin synthase; CHI, chalcone isomerase; CHR, chalcone reductase; CHS, chalcone synthase; DFR, dihydroflavonol 4-reductase; F2H, flavanone 2-hydroxylase; F3H, flavanone 3-hydroxylase; F3'H, flavonoid 3'-hydroxylase; F3'5'H, flavonoid 3'5'-hydroxylase; FLS, flavonol synthase; FNS, flavone synthase; HID, 2-hydroxyisoflavone dehydratase; IFS, isoflavone synthase; LAR, leucoanthocyanidin reductase; LDOX, leucoanthocyanidin dioxygenase.

addition, the genes encoding transcription factors that regulate anthocyanin biosynthesis have evolved more rapidly than the structural genes (Rausher et al., 1999).

THE EVOLUTIONARY HISTORY OF FLAVONOID BIOSYNTHETIC GENES

CHS Is a Representative of the Type III PKS Superfamily

The *CHS* genes are widely distributed in plants, from bryophytes to angiosperms (Jiang et al., 2008; Shimizu et al., 2017; Liou et al., 2018), but they have not been found in other organisms. Every land plant species with available genomic data has at least one putative *CHS* gene (Shimizu et al., 2017). *CHS* is a member of the type III PKS superfamily, which provides diverse polyketide scaffolds of secondary metabolites (Figure 3) (Winkel-Shirley, 2001; Austin and Noel, 2003; Abe and Morita, 2010). The type III PKSs belong to the thiolase superfamily (Jiang et al., 2008). They are found in land plants, microalgae, fungi, and bacteria, but are not found in animals or archaea (Shimizu et al., 2017). The plant type III PKSs retain the overall folded protein structure and the Cys-His-Asn catalytic triad that characterize the *Escherichia coli* 3-ketoacyl-ACP synthase isoform III (KASIII) enzyme, which is involved in *de novo* fatty acid synthesis (Ferrer et al., 1999; Austin and Noel, 2003). The number of land plant type III PKS genes is highly variable among species in the same taxa; for instance, eudicot species have two to 42 genes, whereas the numbers in fungi and bacteria are relatively low (less than five) (Shimizu et al., 2017). The abundance of type III PKSs may contribute the variety of specialized plant metabolites.

The functional diversity of the type III PKSs is generally derived from differences in the starter molecules, the numbers of chain elongation steps, and the mechanisms of the cyclization reactions (Austin and Noel, 2003; Abe and Morita, 2010). The stilbene synthases (STs), which produce stilbenes such as resveratrol, are also type III PKSs. CHSs and STs generate the same intermediates from the same starter molecules using the same chain elongation steps, but they catalyze different intra-molecular cyclizations and produce different products (Austin et al., 2004). The CHSs and STs from the same plant genera are usually classified as the closest neighbors in phylogenetic trees (Figure 3). Data from angiosperms (*Vitis vinifera*, *Arachis hypogaea*, and *Sorghum bicolor*), gymnosperms (*P. sylvestris*), and ferns (*Psilotum nudum*) suggest that after diverging, the STs and CHSs evolved independently (Yu et al., 2005; Weng and Noel, 2012b).

Basal Land Plant CHSs and Non-CHSs

The eudicot *Arabidopsis thaliana* contains four type III PKS genes, including a single functional *CHS* (Kim et al., 2010). In contrast, the bryophyte moss *Physcomitrella patens* has around 19 *CHS* family genes and four other type III PKS genes (Jiang et al., 2006; Wolf et al., 2010). Five of the moss *CHS* family genes were derived from a whole genome duplication, and four are suggested to be derived from segmental duplication and transposition

(Wolf et al., 2010). Four of the genes are upregulated by broadband UV-B irradiation, and the moss plants show increased levels of a flavonol derivative under UV-B illumination. These observations support the hypothesis that the genes and enzymes involved in the UV stress response evolved with the water-to-land transition, when early plants were exposed to increased levels of sunlight (Wolf et al., 2010).

In the bryophyte liverwort *M. polymorpha*, UV-B irradiation and nutrient deprivation significantly increase the total flavone glycoside content (Albert et al., 2018; Clayton et al., 2018). *M. polymorpha* has 24 *CHS* family genes (Bowman et al., 2017) and one of these is significantly upregulated by UV-B stress treatment. This induction was enhanced in transgenic *Marchantia* plants that overexpressed the gene encoding the UV RESISTANCE LOCUS8 (UVR8) photoreceptor (Clayton et al., 2018). Furthermore, transgenic *Marchantia* plants overexpressing a *MpMyb14* transcription factor gene showed increased expression levels of the same *CHS* gene under normal growth conditions (Albert et al., 2018). Knockout *Mpmyb14* mutants were partially impaired by an increase in *CHS* expression levels under nitrogen deficient conditions (Kubo et al., 2018). The results suggest that these liverwort species have at least one *CHS* gene that is activated by the UVR8 signal transduction pathway.

Phylogenetic analyses show close relationships between basal land plant (bryophyte and lycophyte) CHSs and non-CHS type III PKSs (Figure 3) (Wanibuchi et al., 2007; Yu et al., 2018). The non-CHS group includes many enzymes involved in the biosynthesis of secondary metabolites, such as acridone synthases, pyrone synthases, bibenzyl synthases, and *p*-coumaroyltriacetic acid synthases (Winkel-Shirley, 2001). These non-CHS PKS enzymes evolved through repeated gene duplication, mutation, and functional diversification from their ancestral plant enzymes.

Ectopic expression of a *CHS* gene from either the bryophyte *P. patens* or the lycophyte *Selaginella moellendorffii* can partially complement the phenotype of an *A. thaliana* *CHS*-null mutant, *transparent testa 4* (Liou et al., 2018). Crystal structures of CHSs from *P. patens*, *S. moellendorffii*, the monilophyte *Equisetum arvense*, the gymnosperm *Pinus sylvestris*, and the angiosperm *A. thaliana* revealed that the reactivity of the catalytic Cys residue (Cys164 in *M. sativa* CHS2) has changed during the 500 million years of evolution of land plants. The Cys residues in the three recent lineages (monilophytes, gymnosperms, and angiosperms) are present in the thiolate anion form, which gives them stronger nucleophilic power (Liou et al., 2018).

The type III PKSs show broad substrate promiscuity. CHSs do not accept bulky substrates, but the lycophyte *Huperzia serrata* HsPKS1 exhibits remarkable substrate tolerance and catalytic potential (Wanibuchi et al., 2007; Morita et al., 2011). *In vitro*, HsPKS1 produces naringenin chalcone and other polyketides, including aromatic tricyclic pyridoisoindole compounds, which are not found in natural products. A single amino acid replacement in HsPKS1 increases its active-site cavity volume and alters the product chain length and the mechanism of the cyclization reaction. This substrate promiscuity in the type III PKSs provides diverse polyketide scaffolds for the subsequent biosynthesis of secondary metabolites in land plants.

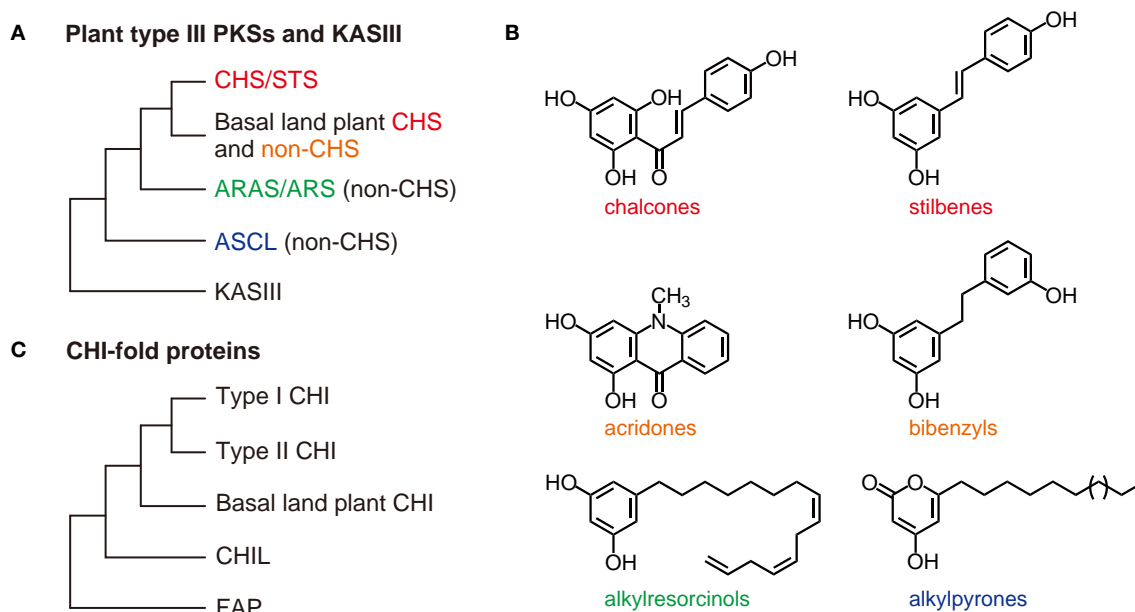


FIGURE 3 | General overview of the type III PKS and CHI-fold protein phylogenies. **(A)** Relationships among the CHS/STS proteins, the basal land plant CHS, and non-CHS proteins of the plant type III PKSs. The non-CHS proteins include the ARAS/ARS proteins and the ASCL families. The overall three-dimensional protein structure is conserved in the type III PKSs and an *E. coli* KASIII enzyme (the $\alpha\beta\alpha\beta$ -fold). **(B)** Examples of type III PKS products. **(C)** The CHI-fold proteins in the CHI, CHIL, and FAP families share a common folded protein structure (the open-faced β -sandwich fold). ARAS, alkylresorcylic acid synthase; ARS, alkylresorcinol synthase; ASCL, anther-specific chalcone synthase-like enzyme; CHIL, CHI-like protein; CHI, chalcone isomerase; CHS, chalcone synthase; FAP, fatty-acid-binding protein; KASIII, 3-ketoacyl-ACP synthase isoform III enzyme; PKSs, polyketide synthases; STS, stilbene synthase.

The ARAS/ARS and ASCL Families

Plant non-CHS type III PKSs also synthesize polyketides from fatty acyl-CoA substrates. Phenolic lipids such as alkylresorcinols are synthesized by alkylresorcinol synthases (ARSs) and alkylresorcylic acid synthases (ARASs) in the monocots *S. bicolor* and *O. sativa*, respectively (Cook et al., 2010; Matsuzawa et al., 2010). Alkylresorcinols in grain crop species show anti-fungal and allelopathic activities.

Hydroxyalkyl- α -pyrone compounds, which are precursors of sporopollenin in the pollen wall exine, are synthesized by anther-specific chalcone synthase-like enzymes (ASCLs) that are specifically and transiently expressed in *A. thaliana* anthers (PKSA/LAP6 and LAP5/PKSB) (Dobritsa et al., 2010; Kim et al., 2010). The *pkas pkbs* double mutant plants are male sterile.

Phylogenetic analyses show that the ARAS/ARS and ASCL families are classified into distinct groups (Figure 3) (Shimizu et al., 2017). Bacteria, fungi, and mosses also have genes encoding type III PKSs that produce these long-chain fatty acyl containing-polyketides (Colpitts et al., 2011; Shimizu et al., 2017; Li et al., 2018). Therefore, these type III PKSs are involved in lipid metabolism across three kingdoms of living organisms.

CHI-Fold Proteins: CHIs and CHILs

The CHIs and CHI-like proteins (CHILs) are members of the CHI-fold family, which also includes the fatty-acid-binding proteins (FAPs) that are involved in fatty acid biosynthesis (Figure 3). These proteins share a common folded three-dimensional structure (Jez et al., 2000; Ngaki et al., 2012; Kaltenbach et al., 2018). The FAP family is distributed in many

bacteria, fungi, and plant species. CHILs partly lack the catalytic amino acid residues conserved in CHIs, but they bind with CHSs and enhance their activity (Ban et al., 2018). Phylogenetic and genomic analyses of CHI-fold proteins suggest that CHILs first appeared in mosses and evolved from FAPs, and then served as the ancestors of CHIs (Ngaki et al., 2012; Morita et al., 2014; Jiang et al., 2015). The CHILs form a group that is distinct from the CHI and FAP groups (Figure 3). The moss *P. patens* has two CHIL and four FAP genes, but does not appear to have any CHI genes (Ngaki et al., 2012; Cheng et al., 2018). The liverwort *M. paleacea* has a CHI, a CHIL, and two FAP genes (Cheng et al., 2018) and the lycophyte *S. moellendorffii* has a CHI, a CHIL, and three FAP genes. *A. thaliana* has a CHI, a CHIL, and three FAP genes whereas the legume *Glycine max* has a type I CHI, three type II CHIs, two CHILs, and six FAP genes (Dastmalchi and Dhaubhadel, 2015; Ban et al., 2018). These results suggest that the number of CHI and CHIL genes remains low in many plant lineages, but that leguminous plants have several CHI genes.

Type I and Type II CHIs in Vascular Plants

There are two types of CHI. Type I CHIs are ubiquitous in vascular plants, whereas type II CHIs are specific to legumes and are involved in flavonoid synthesis during nitrogen-fixing root nodule symbioses (Shimada et al., 2003; Subramanian et al., 2007). Phylogenetic analyses show that the type II CHIs form a distinct group from the type I CHIs (Figure 3). The type II CHIs likely evolved from ancestral CHI-fold proteins (Cheng et al., 2018).

The type II CHIs isomerize both naringenin chalcone and isoliquiritigenin (6'-deoxychalcone) to produce (2S)-naringenin and liquiritigenin (5-deoxyflavanone), respectively. The model legume *Lotus japonicus* has a type I and three type II *CHI* genes (Shimada et al., 2003). These four *CHI* genes form a tandem cluster within a 15-kb region of the genome. In soybean, a type I and two type II *CHI* genes are organized in a gene cluster on chromosome 20 (Dastmalchi and Dhaubhadel, 2015), and are probably derived from tandem gene duplications. The role of the CHIs in legume symbiosis remains unclear. Nitrogen-fixing root nodule symbioses are found in four angiosperm orders: Fabales, Fagales, Cucurbitales, and Rosales (the latter three are known as actinorhizal plants). The expression of a *CHI* gene is upregulated in nodules of the actinorhizal plant *Datisca glomerata* (Gifford et al., 2018), suggesting that these CHIs contribute to increases in flavonoid contents.

CHIs in Basal Land Plants

The *A. thaliana* *chi*-deficient *tt5* mutants produce pale yellow seeds due to significant reductions in proanthocyanidin production. This phenotype is largely complemented by ectopic expression of a *CHI* gene from either the liverwort *M. paleacea* (*MpCHI1*) or the lycophyte *S. moellendorffii* (*SmCHI1*) (Cheng et al., 2018). The CHIs from basal land plants have broad substrate specificities and are more like the type II CHIs than the type I CHIs. Phylogenetic analyses show that *MpCHI1* and *SmCHI1* form separate groups from the type I and type II CHIs (Figure 3) (Cheng et al., 2018; Kaltenbach et al., 2018).

The liverwort *M. polymorpha* also has a single *CHI* gene (Clayton et al., 2018), and *Marchantia chi* mutants do not contain detectable levels of flavone compounds. These *chi* mutants are highly sensitive to UV-B stress treatment. These results indicate that this basal plant species already has a gene encoding a *bona fide* CHI to catalyze the cyclization of naringenin chalcone in the flavonoid biosynthetic pathway.

The CHIL Family

CHILs are categorized as type IV CHI-fold proteins and are found in basal and higher plant species including mosses, liverworts, lycophytes, ferns, gymnosperms, and angiosperms. CHILs do not have *bona fide* CHI activity. However, the RNAi knockdown of *CHIL* expression in *Petunia hybrida* and *Torenia hybrida* resulted in decreased levels of total flavonoids in the flowers (Morita et al., 2014). Three independent *Ipomoea nil* (Japanese morning glory) mutants with alterations in *enhancer of flavonoid production* (a *CHIL* gene) showed pale-colored flower phenotypes (Morita et al., 2014). *CHIL* loss-of-function mutants in *A. thaliana* show reductions in the levels of proanthocyanidin and flavonols in seeds, and flavonols in leaves (Jiang et al., 2015). However, the *A. thaliana* *CHIL* gene could not rescue the phenotypes of the *tt5* mutants (Jiang et al., 2015). The liverwort *M. polymorpha* has a *CHIL* gene (Clayton et al., 2018), and the total flavone content is reduced in *Marchantia chil* mutants under normal growth conditions and under UV-B treatment. Thus, the CHILs enhance total flavonoid production but have roles that are distinct from those of the CHIs. As mentioned above, it was recently shown that CHILs from various plant lineages

(*A. thaliana*, *O. sativa*, *S. moellendorffii*, and *P. patens*) can bind CHSs and boost CHS activity (Ban et al., 2018). CHIs also bind CHSs (Jorgensen et al., 2005), and it may be that the type I and type II CHIs evolved from CHILs and gained their CHI activity during subsequent evolutionary processes (Ban et al., 2018).

Acquisition of Enzymatic CHI Activity During Evolution

The emergence of enzymatic CHIs in plants is an interesting topic in protein evolution, because the CHILs and FAPs are non-enzymatic proteins. The cleft in the CHI active site consists of three highly conserved amino acid residues (Arg36, Thr48, and Tyr106 in the *M. sativa* type II CHI sequence) and their neighboring residues are also conserved (Jez et al., 2000; Ngaki et al., 2012; Kaltenbach et al., 2018). The common ancestral proteins of the CHIs and CHILs were inferred in an extensive phylogenetic analysis, and it appears that all three key catalytic residues were conserved in the ancestral proteins but were inactive (Kaltenbach et al., 2018). The authors performed a stepwise, activity-based screening of recombinant ancestral proteins using an *E. coli* expression system. The results indicated that mutations in amino acid residues other than the catalytic residues were required to initiate CHI evolution, and to acquire CHI catalytic activity.

The 2-Oxoglutarate-Dependent Dioxygenase Family: F3H, FNS I, FLS, and LDOX/ANS

The 2OGD superfamily is one of the largest protein families in the plant kingdom. Its members are widely distributed in bacteria, fungi, plants, vertebrates, and even viruses (van den Born et al., 2008; Farrow and Facchini, 2014; Markolovic et al., 2015; Wu et al., 2016). The 2OGDs are non-heme iron containing enzymes that are localized in the cytosol. 2OGDs incorporate 2-oxoglutarate (2OG or α -ketoglutarate) and activated O₂ into a variety of substrates to form the oxidized products along with succinate and CO₂ ($\text{RH} + 2\text{OG} + \text{O}_2 \rightarrow \text{ROH} + \text{succinate} + \text{CO}_2$). 2OGDs catalyze various oxidative reactions including hydroxylation, halogenation, desaturation, and epimerization (Martinez and Hausinger, 2015) and play important roles in DNA and RNA repair, fatty acid metabolism, oxygen sensing, and biosynthesis of natural products (Farrow and Facchini, 2014; Hagel and Facchini, 2018; Herr and Hausinger, 2018; Islam et al., 2018).

In plants, 2OGDs are involved in histone demethylation, iron sensing, phytohormone metabolism, and the biosynthesis of secondary metabolites (reviewed in Farrow and Facchini, 2014). A phylogenetic analysis of plant 2OGDs from *Chlamydomonas reinhardtii*, *P. patens*, *S. moellendorffii*, *Picea abies*, *O. sativa*, and *A. thaliana* found 3 classes, which the authors named DOXA, DOXB, and DOXC (Kawai et al., 2014). Each class contains proteins from all six species. The DOXA class contains homologs of *E. coli* AlkB; these enzymes are involved in DNA repair (Lindahl et al., 1988; Meza et al., 2012; Mielecki et al., 2012). The DOXB class contains prolyl 4-hydroxylases that catalyze the hydroxylation of proline residues in plant cell

wall proteins (Hieta and Myllyharju, 2002). Proteins in the DOXC class are involved in phytohormone metabolism and the biosynthesis of secondary metabolites including flavonoids, terpenoids, alkaloids, and glucosinolates. The numbers of genes encoding DOXA and DOXB enzymes are limited in the six species, however, the DOXC genes are significantly expanded in the land plants (Kawai et al., 2014).

Four flavonoid biosynthetic enzymes, F3H, FNS I, FLS, and LDOX/ANS belong to the DOXC class. A phylogenetic analysis showed that the genes in the DOXC class can be classified into over 50 clades (Kawai et al., 2014). The *F3H* and *FNS I* genes are in the DOXC28 clade while the *FLS* and *LDOX/ANS* genes are in DOXC47. Among the flavonoid biosynthetic genes in the 2OGD superfamily, it has been proposed that *F3H* was the first to appear (Rauscher, 2006). *FNS I* seems to exist only in the Apiaceae, and it is likely that *FNS I* evolved from *F3H* as a paraphyletic gene (Martens et al., 2003; Gebhardt et al., 2005).

Arabidopsis thaliana contains one *F3H* gene, six *FLS* genes (*AtFLS1*–*AtFLS6*), one *LDOX/ANS* gene, and no *FNS* gene. *AtFLS1* is the major *FLS* (Owens et al., 2008a; Saito et al., 2013). An analysis of structural divergence between duplicated genes showed that transposed duplication (<16 million years ago) explains the relationship between *AtFLS6* (At5g43935) and *AtF3H* (At3g51240) (Wang et al., 2013). Furthermore, the relationship between *AtFLS1* (At5g08640) and *AtFLS5* (At5g63600) is likely explained by a whole genome duplication, and those between *AtFLS2* (At5g63580) and *AtFLS3* (At5g63590), *AtFLS3* and *AtFLS4* (At5g6359580), and *AtFLS4* and *AtFLS5* are most likely explained by tandem local duplications (Wang et al., 2013). The *A. thaliana* *AtLDOX/ANS* has been shown to produce flavonols *in planta* (Preuss et al., 2009). Furthermore, *FLS* and *LDOX/ANS* can partially complement *F3H* function *in vivo*, and this results in the leaky phenotype of *tt6* mutants with null mutations in *AtF3H* (Owens et al., 2008b).

Together, the results suggest that *F3H* is the ancestral 2OGD gene for flavonoid biosynthesis, and that *FLS* and *FNS I* evolved via divergence from *F3H*. No apparent orthologs of either the DOXC28 or DOXC47 clade genes were found in *P. patens* and *S. moellendorffii*, even though these plants produce flavonols (*P. patens*) and flavones (*S. moellendorffii*) (Wolf et al., 2010; Weng and Noel, 2013). However, the liverwort *Plagiochasma appendiculatum* has an active *FNS I* (PaFNS I), and a phylogenetic analysis revealed that PaFNS I is related to the angiosperm *FNS I* and *F3H* proteins, even though it is not in the same clade as them (Han et al., 2014). These data suggest that in *Physcomitrella* and *Selaginella*, the 2OGDs are present in distinct clade(s), or that unrelated enzymes perform the same functions as *F3H*, *FNS*, and/or *FLS*.

The PaFNS I can convert naringenin to either 2-hydroxynaringenin or apigenin (Han et al., 2014). The common horsetail *E. arvense* L also has 2OGD-type *FNS I* activity (Bredebach et al., 2011). Further research in the 2OGD genes of bryophytes, lycophytes, and ferns will help to clarify their evolutionary processes.

The Cytochrome P450 Superfamily: FNSII, F2H, IFS, F3'H, and F3'5'H

The CYPs are widely distributed in viruses, archaea, bacteria, and eukaryotes. They catalyze monooxygenase/hydroxylation reactions in various primary and secondary metabolic processes by insertion of an O atom from molecular O₂ (Mizutani and Ohta, 2010). In eukaryotes, the CYPs are heme-containing membrane proteins localized on the cytosolic surface of the endoplasmic reticulum.

In plants, the CYPs form the largest superfamily of enzymes and account for about 1% of the total number of gene products (Mizutani and Ohta, 2010; Nelson and Werck-Reichhart, 2011; Kawai et al., 2014). The CYPs are categorized into families (e.g., CYP75) that have ~40% or more amino acid sequence identity, and those with 55 % or more identity are categorized into subfamilies (e.g., CYP75A). Furthermore, the plant CYP families (CYP71 to CYP99 and CYP701–) can be classified into clans whose members are derived from single ancestors. The land plant CYPs form 11 clans and seven of these (clans 51, 74, 97, 710, 711, 727, and 746) consist of single CYP families, while the remaining four (clans 71, 72, 85, and 86) include proteins in multiple CYP families (Nelson and Werck-Reichhart, 2011). Green algae contain CYPs in five single family clans (clans 51, 97, 710, 711, and 746), and members of these clans are involved in fundamental biological processes such as biosynthesis of sterols, xanthophylls, and phytohormones (Nelson, 2006; Nelson and Werck-Reichhart, 2011). Therefore, it is likely that these clans include the ancestral CYPs. The multi-family CYP clans have become highly diversified during plant evolution (Nelson and Werck-Reichhart, 2011). Some CYP families in these clans (clans 71, 72, 85, and 86) are present in bryophytes and/or liverworts but not in green algae. Additional novel CYP families were gained in stepwise processes following the evolution of vascular plants.

Flavonoid biosynthetic enzymes (members of the CYP75 and CYP93 families) are classified into clan 71. This clan contains the largest number of CYP families, and in addition to the flavonoid biosynthetic enzymes, it includes families involved in the biosynthesis of phenylpropanoids (CYP73, CYP84, CYP98), alkaloids (CYP80, CYP82, CYP719), terpenoids (CYP76, CYP99, CYP705, CYP706, CYP726), and glucosinolates (CYP79, CYP83). The clan 71 families CYP73, CYP74, CYP78, CYP88, CYP98, CYP701, CYP703, CYP736, and CYP761 are present in bryophytes, suggesting that these may be ancestral families.

The CYP75 Family: F3'H and F3'5'H

The F3'H and F3'5'H enzymes generally belong to the CYP75B and CYP75A families, respectively. The CYP75 family enzymes from monocots and dicots form distinct clusters within each subfamily, suggesting that the F3'H and F3'5'H functions were established before the divergence of monocots and dicots. However, there are some exceptions. In the Asteraceae, F3'5'H belongs to the CYP75B subfamily and forms a distinct cluster from F3'H (Seitz et al., 2006). This suggests that the Asteraceae F3'H gained F3'5'H activity before speciation but after the separation of the monocots and dicots. Similarly, the rice

CYP75B4 catalyzes the 5'-hydroxylation of 3'-methoxyflavone chrysoeriol, and also functions as an F3'H (Lam et al., 2015). Rice contains CYP75B3 as an F3'H and CYP75A11 as a non-functional F3'5'H. Such CYP gene distributions in the CYP75A and CYP75B subfamilies are also found in other Poaceae plants. These data suggest that the CYP75A and CYP75B subfamilies separated before divergence of monocots and dicots, and that genes in the CYP75B subfamily later gained F3'5'H activity, at least in the Asteraceae and Poaceae.

Arabidopsis thaliana has a single gene for F3'H (CYP75B1, At5g07990) and no genes corresponding to F3'5'H, FNS II, or F2H. The CYP75B1 gene appears to be related to the CYP701A (At5g25900) gene, which encodes an enzyme involved in gibberellin biosynthesis, via a transposed duplication that occurred 16–107 million years ago and it was proposed that CYP75B1 is the parental locus of CYP701A (Wang et al., 2013). However, the CYP701 family is distributed among bryophytes and vascular plants while the CYP75 family is found in gymnosperms and angiosperms. In addition, a phylogenetic analysis suggested that the moss/liverwort-specific CYP761 family is closely related to the CYP75 family (Nelson and Werck-Reichhart, 2011). These data suggest that either CYP701 or CYP761 is the ancestral family of CYP75. Although 3'-hydroxylated flavone derivatives were detected in *M. polymorpha* and *S. moellendorffii* (Markham et al., 1998), CYP75 family members can be found only in gymnosperms and angiosperms (Nelson and Werck-Reichhart, 2011). Therefore, in bryophytes, lycophytes, and ferns, enzymes from other CYP families may function as F3'H and/or F3'5'H.

Such inconsistencies in the relationships between metabolites and genes are also observed in phenylpropanoid metabolism. Among the CYPs involved in phenylpropanoid metabolism from clan 71, CYP73A (cinnamate 4-hydroxylase, C4H), CYP98A (*p*-coumaroyl shikimate 3'-hydroxylase, C3'H), CYP73, and CYP98 first appeared in liverworts and mosses, while CYP84 (ferulate 5-hydroxylase, F5H) is found only in angiosperms (Nelson and Werck-Reichhart, 2011). However, the syringyl lignin units from the phenylpropanoid pathway are distributed in land plants (Bowman et al., 2017), suggesting that genes in other CYP families may have F5H activities. In fact, the *Selaginella* CYP788A1 functions as a F5H (Weng et al., 2008).

The CYP93 Family: FNS II, F2H, and IFS

A genome-wide analysis of the CYP93 family genes from 60 green plants indicated that the CYP93 family is found only in angiosperms (Du et al., 2016). Among the 10 subfamilies (CYP93A–CYP93K), CYP93A is the ancestral group distributed in both monocots and dicots; CYP93B and CYP93C are distributed only in dicots; CYP93G and CYP93J are found only in monocots; and CYP93E and CYP93F are specific to legumes and grasses, respectively. Thus, the CYP93 family shows plant lineage-specific evolution.

The CYP93 family contains enzymes involved in the biosynthesis of flavones and isoflavones (FNS II, F2H, and IFS). The monocot FNS II and F2H belong to the CYP93G subfamily, while those in the dicots are categorized in the CYP93B subfamily. Therefore, phylogenetic analyses suggest that the functions of the FNSII and F2H enzymes were established

after the divergence of monocots and dicots. The IFS enzyme in legumes is a member of the CYP93C subfamily, which may be derived from the CYP93B subfamily (Du et al., 2016).

As with the CYP75 family, there are inconsistencies in the relationships between metabolites and genes in the CYP93 family. The CYP93 family genes are found only in angiosperms (Nelson and Werck-Reichhart, 2011), but flavones are widely distributed in plants, from bryophytes to angiosperms (**Figure 1**). This suggests that enzymes involved in flavone biosynthesis belong to the CYP93 family in some species of angiosperms, but that different CYP families and/or other enzymes play the same roles in other plant taxa. For example, the Apiaceae family adapted the 2OGD-type FNS I rather than the CYP-type FNS II to produce flavones (Gebhardt et al., 2007).

Isoflavonoids are the typical flavonoids found in legumes, and isoflavone biosynthetic genes are found only in legumes. However, isoflavonoids are also found in non-legume plants including *Iris*, which contains a wide variety of isoflavones, some mosses (e.g., *Bryum capillare*), gymnosperms, and monocot and dicot angiosperms (Dewick, 1994; Lapcik, 2007). This suggests that plants independently acquired the ability to produce isoflavonoids during the evolution of the CYP families, except for CYP93C and/or other non-CYP enzymes that function as IFSs.

The Short-Chain Dehydrogenase/Reductase Family: DFR, ANR, and LAR

DFR, ANR, and LAR are members of the SDR superfamily, which is widely distributed in viruses, archaea, prokaryotes, and eukaryotes (Jornvall et al., 1999; Kavanagh et al., 2008). The SDRs constitute one of the largest NAD(P)(H) dependent oxidoreductase families and are involved in the primary metabolism of lipids, carbohydrates, and hormones, and the secondary metabolism of molecules such as terpenoids, alkaloids, and phenolic compounds (Jornvall et al., 1999; Kavanagh et al., 2008; Tonfack et al., 2011). In spite of the low overall sequence similarities among SDRs (15–30 %), the SDRs possess a conserved 3D structure consisting of a Rossmann-fold β -sheet surrounded by α -helices for nucleotide binding. Generally, the SDRs can be classified into several types ("classical," "extended," "intermediate," "divergent," and "complex") based on their primary structures, cofactor binding motifs, and active sites (Kavanagh et al., 2008; Moummou et al., 2012). A recent study revealed that the "intermediate" and "complex" types are not found in plants while the types "atypical" and "unknown" are found in plants. Thus, the plant SDRs can be categorized into five types: "classical," "divergent," "extended," "atypical," and "unknown" (Moummou et al., 2012). Among the plant SDRs, "classical" and "extended" are the major types, as they are in other organisms. The "classical" type is composed of about 250 amino acid residues, and the "extended" type has a domain of 100 additional amino acid residues in the C-terminal region. The "atypical" type, an uncommon type of SDRs, was included in the SDR family because of its Rossmann-fold structure, which is typical of SDRs (Moummou et al., 2012).

A genome sequence analysis using 10 species including *C. reinhardtii*, *P. patens*, *S. moellendorffii*, four dicots (*A. thaliana*,

Populus trichocarpa, *V. vinifera*, and *G. max*) and three monocots (*O. sativa*, *S. bicolor*, and *Z. mays*) showed that most plant SDRs can be classified into 49 families that are distributed among the five types mentioned above (Moummou et al., 2012). The DFR and ANR enzymes are classified into the SDR108E family in the “extended” type. The LAR enzymes belong to the SDR460A family, which is an “atypical” type (Kallberg et al., 2010; Moummou et al., 2012).

The SDR108E Family: DFR and ANR

Compared to other families that contain a few genes per species, the SDR108E family contains the largest number of SDR genes; for example, this family includes 24 genes in *A. thaliana* and 44 genes in *O. sativa*. Furthermore, the SDR108E family shows the lowest average sequence identity, indicating that the SDR108E family genes are highly diversified (Moummou et al., 2012). The distribution of SDR108E family genes in 10 species indicates that this family has expanded significantly in vascular plants.

In addition to the DFR and ANR enzymes, the SDR108E family contains other enzymes involved in secondary metabolism, including cinnamoyl-CoA reductase (CCR) for lignin biosynthesis, and phenylacetaldehyde reductase for the production of volatile 2-phenylethanol. Furthermore, enzymes involved in phytohormone metabolism, such as phaseic acid reductase for abscisic acid catabolism (Weng et al., 2016) and BEN1 for brassinosteroid homeostasis (Yuan et al., 2007) belong to this family. Each enzyme type forms a distinct cluster in the phylogenetic tree. The CCR and phenylacetaldehyde reductase clusters contain enzymes from *P. patens* and *S. moellendorffii*, whereas the clusters of DFRs, ANRs, phaseic acid reductases, and BEN1 enzymes are derived from flowering plants. These data suggest that these DFRs, ANRs, and other enzymes appeared more recently than the CCRs and phenylacetaldehyde reductases.

Other SDR families show similar gene expansion patterns to that of SDR108E. These families (SDR110C, SDR114C, SDR65C, and SDR460A) contain genes involved in the biosynthesis of alkaloids, terpenoids, phenylpropanoids, and phytohormones (Moummou et al., 2012). SDR families that are less diversified contain genes involved in primary metabolism, such as lipid and chlorophyll biosynthesis (Moummou et al., 2012).

The SDR460A Family: LAR

The SDR460A family is also referred to as the PIP family, named after the first three enzymes (pinoresinol-laricirecinol reductase, isoflavone reductase, and phenylcoumaran benzylic ether reductase) that were discovered to belong to this family (Gang et al., 1999; Min et al., 2003; Wang et al., 2006). In addition, vestitone reductase, eugenol synthase, and isoeugenol synthase also belong to this family (Koeduka et al., 2008). Pinoresinol-laricirecinol reductase and phenylcoumaran benzylic ether reductase function in the lignan biosynthetic pathway; isoflavone reductase and vestitone reductase are involved in isoflavonoid biosynthesis; and eugenol synthase and isoeugenol synthase are involved in the biosynthesis of volatile phenylpropenes. Thus, the SDR460A family members are involved in the biosynthesis of various phenolic compounds. The SDR460A family members (“atypical” type SDRs) are limited to and has greatly expanded in

vascular plants (Moummou et al., 2012), suggesting that they are needed for vascular plant prosperity.

A phylogenetic analysis of LARs from various plants revealed that the plant LARs can be classified into two clusters: proteins from dicotyledons, and proteins from monocotyledons and gymnosperms (Wang et al., 2018). Therefore, the monocotyledon LARs are more closely related to the gymnosperm LARS than the dicotyledon LARS.

FUTURE PERSPECTIVES

The progress in metabolomics technologies including chemoinformatics and the abundant genomic information of flavonoid biosynthetic genes facilitated a fully understanding of evolution of the flavonoid/phenylpropanoid metabolisms in plant kingdom. In this review, we have focused on the evolution of enzymes involved in the biosynthesis of flavonoid skeleton molecules. The basic structures of flavonoids are formed by a type III PKS, CHS. The broad substrate promiscuity and functional diversity of the type III PKSs may be a driving force for expanding the chemical variety of specialized metabolites. In other specialized metabolisms, isomerases (e.g., oxidosqualene synthase) and lyases (e.g., terpene synthase) are also involved in scaffold formation and may contribute the chemical diversity of secondary metabolites. Modification enzymes such as GTs (glycosyltransferases) and acyltransferases also contribute greatly to the huge diversity of flavonoids and other secondary metabolites. Interestingly, plants have two types of flavonoid GTs: the cytosolic family 1 GTs and the vacuolar glycoside hydrolase family 1 (GH1) (Cao et al., 2017). Acylation is also catalyzed by differentially localized enzymes: cytosolic BAHD acyltransferases and vacuolar SCPL acyltransferases, derived from serine carboxypeptidase (Milkowski and Strack, 2004; Moghe and Last, 2015). It is still unknown why plants have evolved these differentially localized enzymes for the modification of flavonoids and other specialized metabolites. The evolution of the family 1 GTs, the GH1s, and the BAHD acyltransferases in plants have been reviewed elsewhere (St-Pierre and De Luca, 2000; Yu et al., 2009; Tuominen et al., 2011; Yonekura-Sakakibara and Hanada, 2011; Caputi et al., 2012; Moghe and Last, 2015; Cao et al., 2017). Throughout their long history, plants have engineered their metabolic pathways to adapt themselves to their habitats and growth conditions in tissue and organ specific manners. A detailed understanding of the evolutionary history of metabolic enzymes involved in biosynthesis, modification, transport, secretion, transcriptional regulation, and chemodiversity will assist us in the engineering of specialized metabolic pathways to produce desirable metabolites with minimal energy expenditures.

AUTHOR CONTRIBUTIONS

KY-S proposed the concept. KY-S, YH, and RN developed it and wrote the manuscript.

FUNDING

This work was partially supported by the JSPS KAKENHI program (grant number 17K07460 to KY-S),

REFERENCES

- Abe, I., and Morita, H. (2010). Structure and function of the chalcone synthase superfamily of plant type iii polyketide synthases. *Nat. Prod. Rep.* 27, 809–838. doi: 10.1039/b909988n
- Afendi, F. M., Okada, T., Yamazaki, M., Hirai-Morita, A., Nakamura, Y., Nakamura, K., et al. (2012). Knapsack family databases: integrated metabolite-plant species databases for multifaceted plant research. *Plant Cell Physiol.* 53:e1. doi: 10.1093/pcp/pcr165
- Akashi, T., Aoki, T., and Ayabe, S. (1998). Identification of a cytochrome p450 cDNA encoding (2s)-flavanone 2-hydroxylase of licorice (*glycyrrhiza echinata* L.; *Fabaceae*) which represents licodione synthase and flavone synthase ii. *FEBS Lett.* 431, 287–290. doi: 10.1016/S0014-5793(98)00781-9
- Akashi, T., Aoki, T., and Ayabe, S. (2005). Molecular and biochemical characterization of 2-hydroxyisoflavanone dehydratase. Involvement of carboxylesterase-like proteins in leguminous isoflavone biosynthesis. *Plant Physiol.* 137, 882–891. doi: 10.1104/pp.104.056747
- Akimoto, N., Ara, T., Nakajima, D., Suda, K., Ikeda, C., Takahashi, S., et al. (2017). Flavonoidsearch: a system for comprehensive flavonoid annotation by mass spectrometry. *Sci. Rep.* 7:1243. doi: 10.1038/s41598-017-01390-3
- Albert, N. W., Thrimawithana, A. H., McGhie, T. K., Clayton, W. A., Deroles, S. C., Schwinn, K. E., et al. (2018). Genetic analysis of the liverwort *Marchantia polymorpha* reveals that r2r3myb activation of flavonoid production in response to abiotic stress is an ancient character in land plants. *New Phytol.* 218, 554–566. doi: 10.1111/nph.15002
- Anderson, Ø. M., and Markham, K. R. (2006). *Flavonoids: Chemistry, Biochemistry and Applications*. Boca Raton: CRC Press.
- Asakawa, Y., Takikawa, K., Toyota, M., and Takemoto, T. (1982). Chemosystematics of the bryophytes.12. Novel bibenzyl derivatives and ent-cuparene-type sesquiterpenoids from radula species. *Phytochemistry* 21, 2481–2490. doi: 10.1016/0031-9422(82)85245-X
- Asakawa, Y., Toyota, M., and Takemoto, T. (1978). 7 new bibenzyls and a dihydrochalcone from radula-variabilis. *Phytochemistry* 17, 2005–2010. doi: 10.1016/S0031-9422(00)88752-X
- Austin, M. B., Bowman, M. E., Ferrer, J. L., Schroder, J., and Noel, J. P. (2004). An aldol switch discovered in stilbene synthases mediates cyclization specificity of type iii polyketide synthases. *Chem. Biol.* 11, 1179–1194. doi: 10.1016/j.chembiol.2004.05.024
- Austin, M. B., and Noel, J. P. (2003). The chalcone synthase superfamily of type iii polyketide synthases. *Nat. Prod. Rep.* 20, 79–110. doi: 10.1039/b100917f
- Ban, Z. N., Qin, H., Mitchell, A. J., Liu, B. X., Zhang, F. X., Weng, J. K., et al. (2018). Noncatalytic chalcone isomerase-fold proteins in humulus lupulus are auxiliary components in prenylated flavonoid biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* 115, E5223–E5232. doi: 10.1073/pnas.1802223115
- Barron, D., and Ibrahim, R. K. (1996). Isoprenylated flavonoids - a survey. *Phytochemistry* 43, 921–982. doi: 10.1016/S0031-9422(96)00344-5
- Bowman, J. L., Kohchi, T., Yamato, K. T., Jenkins, J., Shu, S. Q., Ishizaki, K., et al. (2017). Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. *Cell* 171, 287–304. doi: 10.1016/j.cell.2017.09.030
- Bredebach, M., Matern, U., and Martens, S. (2011). Three 2-oxoglutarate-dependent dioxygenase activities of equisetum arvense L. Forming flavone and flavonol from (2s)-naringenin. *Phytochemistry* 72, 557–563. doi: 10.1016/j.phytochem.2011.01.036
- Cao, Y. Y., Yang, J. F., Liu, T. Y., Su, Z. F., Zhu, F. Y., Chen, M. X., et al. (2017). A phylogenetically informed comparison of gh1 hydrolases between *Arabidopsis* and rice response to stressors. *Front. Plant Sci.* 8:350. doi: 10.3389/fpls.2017.00350
- Caputi, L., Malnoy, M., Goremykin, V., Nikiforova, S., and Martens, S. (2012). A genome-wide phylogenetic reconstruction of family 1 udp-glycosyltransferases revealed the expansion of the family during the adaptation of plants to life on land. *Plant J.* 69, 1030–1042. doi: 10.1111/j.1365-313X.2011.04853.x
- Carrington, Y., Guo, J., Le, C. H., Fillo, A., Kwon, J., Tran, L. T., et al. (2018). Evolution of a secondary metabolic pathway from primary metabolism: shikimate and quinate biosynthesis in plants. *Plant J.* 95, 823–833. doi: 10.1111/tpj.13990
- Cheng, A. X., Zhang, X. B., Han, X. J., Zhang, Y. Y., Gao, S., Liu, C. J., et al. (2018). Identification of chalcone isomerase in the basal land plants reveals an ancient evolution of enzymatic cyclization activity for synthesis of flavonoids. *New Phytol.* 217, 909–924. doi: 10.1111/nph.14852
- Clayton, W. A., Albert, N. W., Thrimawithana, A. H., McGhie, T. K., Deroles, S. C., Schwinn, K. E., et al. (2018). Uvr8-mediated induction of flavonoid biosynthesis for uvb tolerance is conserved between the liverwort *Marchantia polymorpha* and flowering plants. *Plant J.* 96, 503–517. doi: 10.1111/tpj.14044
- Colpitts, C. C., Kim, S. S., Posehn, S. E., Jepson, C., Kim, S. Y., Wiedemann, G., et al. (2011). Ppscl, a moss ortholog of anther-specific chalcone synthase-like enzymes, is a hydroxyalkylpyrone synthase involved in an evolutionarily conserved sporopollenin biosynthesis pathway. *New Phytol.* 192, 855–868. doi: 10.1111/j.1469-8137.2011.03858.x
- Cook, D., Rimando, A. M., Clemente, T. E., Schroder, J., Dayan, F. E., Nanayakkara, N. P. D., et al. (2010). Alkylresorcinol synthases expressed in sorghum bicolor root hairs play an essential role in the biosynthesis of the allelopathic benzoquinone sorgoleone. *Plant Cell.* 22, 867–887. doi: 10.1105/tpc.109.072397
- Dastmalchi, M., and Dhaubhadel, S. (2015). Soybean chalcone isomerase: evolution of the fold, and the differential expression and localization of the gene family. *Planta* 241, 507–523. doi: 10.1007/s00425-014-2200-5
- Dewick, P. M. (1994). "Isoflavonoids," in *The Flavonoids Advances in Research Since 1986*, ed J. B. Harborne (Chapman and Hall/CRC: United States of America), 117–238.
- Dobritsa, A. A., Lei, Z. T., Nishikawa, S., Urbanczyk-Wochniak, E., Huhman, D. V., Preuss, D., et al. (2010). Lap5 and lap6 encode anther-specific proteins with similarity to chalcone synthase essential for pollen exine development in *Arabidopsis*. *Plant Physiol.* 153, 937–955. doi: 10.1104/pp.110.157446
- Du, H., Ran, F., Dong, H. L., Wen, J., Li, J. N., and Liang, Z. (2016). Genome-wide analysis, classification, evolution, and expression analysis of the cytochrome p450 93 family in land plants. *PLoS ONE* 11:e0165020. doi: 10.1371/journal.pone.0165020
- Farrow, S. C., and Facchini, P. J. (2014). Functional diversity of 2-oxoglutarate/fe(ii)-dependent dioxygenases in plant metabolism. *Front. Plant Sci.* 5:524. doi: 10.3389/fpls.2014.00524
- Ferrer, J. L., Jez, J. M., Bowman, M. E., Dixon, R. A., and Noel, J. P. (1999). Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. *Nat. Struct. Biol.* 6, 775–784. doi: 10.1038/11553
- Gang, D. R., Kasahara, H., Xia, Z. Q., Vander Mijnsbrugge, K., Bauw, G., Boerjan, W., et al. (1999). Evolution of plant defense mechanisms. Relationships of phenylcoumaran benzylic ether reductases to pinorensinol-laricresinol and isoflavone reductases. *J. Biol. Chem.* 274, 7516–7527. doi: 10.1074/jbc.274.11.7516
- Gebhardt, Y., Witte, S., Forkmann, G., Lukacin, R., Matern, U., and Martens, S. (2005). Molecular evolution of flavonoid dioxygenases in the family apiaceae. *Phytochemistry* 66, 1273–1284. doi: 10.1016/j.phytochem.2005.03.030

- Gebhardt, Y. H., Witte, S., Steuber, H., Matern, U., and Martens, S. (2007). Evolution of flavone synthase i from parsley flavanone 3 β -hydroxylase by site-directed mutagenesis. *Plant Physiol.* 144, 1442–1454. doi: 10.1104/pp.107.098392
- Gifford, I., Battenberg, K., Vaniya, A., Wilson, A., Tian, L., Fiehn, O., et al. (2018). Distinctive patterns of flavonoid biosynthesis in roots and nodules of *Datisca glomerata* and *medicago* spp. Revealed by metabolomic and gene expression profiles. *Front. Plant Sci.* 9:1463. doi: 10.3389/fpls.2018.01463
- Goiris, K., Muylaert, K., Voorspoels, S., Noten, B., De Paepe, D., Baart, G. J. E., et al. (2014). Detection of flavonoids in microalgae from different evolutionary lineages. *J. Phycol.* 50, 483–492. doi: 10.1111/jpy.12180
- Gould, K. S., and Lister, C. (2006). “Flavonoid functions in plants,” in *Flavonoids: Chemistry, Biochemistry and Applications*, eds Ø. M. Anderson and K. R. Markham (Boca Raton: CRC Press), 397–442.
- Hagel, J. M., and Facchini, P. J. (2018). Expanding the roles for 2-oxoglutarate-dependent oxygenases in plant metabolism. *Nat. Prod. Rep.* 35, 721–734. doi: 10.1039/C7NP00060J
- Han, X. J., Wu, Y. F., Gao, S., Yu, H. N., Xu, R. X., Lou, H. X., et al. (2014). Functional characterization of a *plagiochasma appendiculatum* flavone synthase i showing flavanone 2-hydroxylase activity. *FEBS Lett.* 588, 2307–2314. doi: 10.1016/j.febslet.2014.05.023
- Harborne, J. B. (1988). *The Flavonoids. Advances in Research Since 1980*. Boston, MA: Springer US.
- Herr, C. Q., and Hausinger, R. P. (2018). Amazing diversity in biochemical roles of Fe(II)/2-oxoglutarate oxygenases. *Trends Biochem. Sci.* 43, 517–532. doi: 10.1016/j.tibs.2018.04.002
- Hietä, R., and Myllyharju, J. (2002). Cloning and characterization of a low molecular weight prolyl 4-hydroxylase from *Arabidopsis thaliana*. Effective hydroxylation of proline-rich, collagen-like, and hypoxia-inducible transcription factor alpha-like peptides. *J. Biol. Chem.* 277, 23965–23971. doi: 10.1074/jbc.M201865200
- Islam, M. S., Leissing, T. M., Chowdhury, R., Hopkinson, R. J., and Schofield, C. J. (2018). 2-oxoglutarate-dependent oxygenases. *Annu. Rev. Biochem.* 87, 585–620. doi: 10.1146/annurev-biochem-061516-044724
- Jez, J. M., Bowman, M. E., Dixon, R. A., and Noel, J. P. (2000). Structure and mechanism of the evolutionarily unique plant enzyme chalcone isomerase. *Nat. Struct. Biol.* 7, 786–791. doi: 10.1038/79025
- Jiang, C. G., Kim, S. Y., and Suh, D. Y. (2008). Divergent evolution of the thiolase superfamily and chalcone synthase family. *Mol. Phylog. Evol.* 49, 691–701. doi: 10.1016/j.ympev.2008.09.002
- Jiang, C. G., Schommer, C. K., Kim, S. Y., and Suh, D. Y. (2006). Cloning and characterization of chalcone synthase from the moss, *Physcomitrella patens*. *Phytochemistry* 67, 2531–2540. doi: 10.1016/j.phytochem.2006.09.030
- Jiang, W. B., Yin, Q. G., Wu, R. R., Zheng, G. S., Liu, J. Y., Dixon, R. A., et al. (2015). Role of a chalcone isomerase-like protein in flavonoid biosynthesis in *Arabidopsis thaliana*. *J. Exp. Bot.* 66, 7165–7179. doi: 10.1093/jxb/erv413
- Jorgensen, K., Rasmussen, A. V., Morant, M., Nielsen, A. H., Bjarnholt, N., Zagrobelny, M., et al. (2005). Metabolite formation and metabolic channeling in the biosynthesis of plant natural products. *Curr. Opin. Plant Biol.* 8, 280–291. doi: 10.1016/j.pbi.2005.03.014
- Jornvall, H., Hoog, J. O., and Persson, B. (1999). Sdr and mdr: completed genome sequences show these protein families to be large, of old origin, and of complex nature. *FEBS Lett.* 445, 261–264. doi: 10.1016/S0014-5793(99)00130-1
- Kallberg, Y., Oppermann, U., and Persson, B. (2010). Classification of the short-chain dehydrogenase/reductase superfamily using hidden Markov models. *FEBS J.* 277, 2375–2386. doi: 10.1111/j.1742-4658.2010.07656.x
- Kaltenbach, M., Burke, J. R., Dindo, M., Pabis, A., Munsberg, F. S., Rabin, A., et al. (2018). Evolution of chalcone isomerase from a noncatalytic ancestor. *Nat. Chem. Biol.* 14, 548–555. doi: 10.1038/s41589-018-0042-3
- Kavanagh, K. L., Jornvall, H., Persson, B., and Oppermann, U. (2008). Medium- and short-chain dehydrogenase/reductase gene and protein families: the sdr superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes. *Cell. Mol. Life Sci.* 65, 3895–3906. doi: 10.1007/s00018-008-8588-y
- Kawai, Y., Ono, E., and Mizutani, M. (2014). Evolution and diversity of the 2-oxoglutarate-dependent dioxygenase superfamily in plants. *Plant J.* 78, 328–343. doi: 10.1111/tpj.12479
- Kim, S. S., Grienemberger, E., Lallemand, B., Colpitts, C. C., Kim, S. Y., Souza, C. D., et al. (2010). Lap6/polyketide synthase a and lap5/polyketide synthase b encode hydroxyalkyl alpha-pyrone synthases required for pollen development and sporopollenin biosynthesis in *Arabidopsis thaliana*. *Plant Cell* 22, 4045–4066. doi: 10.1105/tpc.110.080028
- Koeduka, T., Louie, G. V., Orlova, I., Kish, C. M., Ibdah, M., Wilkerson, C. G., et al. (2008). The multiple phenylpropene synthases in both *Clarkia breweri* and *Petunia hybrida* represent two distinct protein lineages. *Plant J.* 54, 362–374. doi: 10.1111/j.1365-3113.2008.03412.x
- Kubo, H., Nozawa, S., Hiwatashi, T., Kondou, Y., Nakabayashi, R., Mori, T., et al. (2018). Biosynthesis of riccionidins and marchantins is regulated by r2r3-myb transcription factors in *Marchantia polymorpha*. *J. Plant Res.* 131, 849–864. doi: 10.1007/s10265-018-1044-7
- Lam, P. Y., Liu, H., and Lo, C. (2015). Completion of tricin biosynthesis pathway in rice: cytochrome p450 75b4 is a unique chrysoeriol 5'-hydroxylase. *Plant Physiol.* 168, 1527–1536. doi: 10.1104/pp.15.00566
- Lam, P. Y., Tobimatsu, Y., Takeda, Y., Suzuki, S., Yamamura, M., Umezawa, T., et al. (2017). Disrupting flavone synthase ii alters lignin and improves biomass digestibility. *Plant Physiol.* 174, 972–985. doi: 10.1104/pp.16.01973
- Lan, W., Lu, F. C., Regner, M., Zhu, Y. M., Rencoret, J., Ralph, S. A., et al. (2015). Tricin, a flavonoid monomer in monocot lignification. *Plant Physiol.* 167, 1284–1295. doi: 10.1104/pp.114.253757
- Lan, W., Morreel, K., Lu, F. C., Rencoret, J., del Rio, J. C., Voorend, W., et al. (2016a). Maize tricin-oligolignol metabolites and their implications for monocot lignification. *Plant Physiol.* 171, 810–820. doi: 10.1104/pp.16.02012
- Lan, W., Rencoret, J., Lu, F. C., Karlen, S. D., Smith, B. G., Harris, P. J., et al. (2016b). Tricin-lignins: occurrence and quantitation of tricin in relation to phylogeny. *Plant J.* 88, 1046–1057. doi: 10.1111/tpj.13315
- Lapcik, O. (2007). Isoflavonoids in non-leguminous taxa: a rarity or a rule? *Phytochemistry* 68, 2909–2916. doi: 10.1016/j.phytochem.2007.08.006
- Li, L., Aslam, M., Rabbi, F., Vanderwel, M. C., Ashton, N. W., and Suh, D. Y. (2018). Ppors, an ancient type iii polyketide synthase, is required for integrity of leaf cuticle and resistance to dehydration in the moss, *Physcomitrella patens*. *Planta* 247, 527–541. doi: 10.1007/s00425-017-2806-5
- Lindahl, T., Sedgwick, B., Sekiguchi, M., and Nakabeppu, Y. (1988). Regulation and expression of the adaptive response to alkylating agents. *Annu. Rev. Biochem.* 57, 133–157. doi: 10.1146/annurev.bi.57.070188.001025
- Liou, G., Chiang, Y. C., Wang, Y., and Weng, J. K. (2018). Mechanistic basis for the evolution of chalcone synthase catalytic cysteine reactivity in land plants. *J. Biol. Chem.* 293, 18601–18612. doi: 10.1074/jbc.RA118.005695
- Markham, K. R., Ryan, K. G., Bloor, S., and Mitchell, K. A. (1998). An increase in the luteolin: apigenin ratio in *Marchantia polymorpha* on UV-B enhancement. *Phytochemistry* 48, 791–794. doi: 10.1016/S0031-9422(97)00875-3
- Markolovic, S., Wilkins, S. E., and Schofield, C. J. (2015). Protein hydroxylation catalyzed by 2-oxoglutarate-dependent oxygenases. *J. Biol. Chem.* 290, 20712–20722. doi: 10.1074/jbc.R115.662627
- Martens, S., Forkmann, G., Britsch, L., Wellmann, F., Matern, U., and Lukacin, R. (2003). Divergent evolution of flavonoid 2-oxoglutarate-dependent dioxygenases in parsley. *FEBS Lett.* 544, 93–98. doi: 10.1016/S0014-5793(03)00479-4
- Martens, S., and Mithofer, A. (2005). Flavones and flavone synthases. *Phytochemistry* 66, 2399–2407. doi: 10.1016/j.phytochem.2005.07.013
- Martinez, S., and Hausinger, R. P. (2015). Catalytic mechanisms of Fe(II)- and 2-oxoglutarate-dependent oxygenases. *J. Biol. Chem.* 290, 20702–20711. doi: 10.1074/jbc.R115.648691
- Matsuzawa, M., Katsuyama, Y., Funai, N., and Horinouchi, S. (2010). Alkylresorcylic acid synthesis by type iii polyketide synthases from rice *Oryza sativa*. *Phytochemistry* 71, 1059–1067. doi: 10.1016/j.phytochem.2010.02.012
- Meza, T. J., Moen, M. N., Vagbo, C. B., Krokan, H. E., Klungland, A., Grini, P. E., et al. (2012). The DNA dioxygenase AlkB2 protects *Arabidopsis thaliana* against methylation damage. *Nucleic Acids Res.* 40, 6620–6631. doi: 10.1093/nar/gks327
- Mielecki, D., Zugaj, D. L., Muszewska, A., Piwowarski, J., Chojnacka, A., Mielecki, M., et al. (2012). Novel AlkB dioxygenases—alternative models for *in silico* and *in vivo* studies. *PLoS ONE* 7:e30588. doi: 10.1371/journal.pone.0030588
- Milkowski, C., and Strack, D. (2004). Serine carboxypeptidase-like acyltransferases. *Phytochemistry* 65, 517–524. doi: 10.1016/j.phytochem.2003.12.018

- Min, T., Kasahara, H., Bedgar, D. L., Youn, B., Lawrence, P. K., Gang, D. R., et al. (2003). Crystal structures of pinoresinol-lariciresinol and phenylcoumaran benzylic ether reductases and their relationship to isoflavone reductases. *J. Biol. Chem.* 278, 50714–50723. doi: 10.1074/jbc.M308493200
- Mizutani, M., and Ohta, D. (2010). Diversification of p450 genes during land plant evolution. *Annu. Rev. Plant Biol.* 61, 291–315. doi: 10.1146/annurev-arplant-042809-112305
- Moghe, G. D., and Last, R. L. (2015). Something old, something new: conserved enzymes and the evolution of novelty in plant specialized metabolism. *Plant Physiol.* 169, 1512–1523. doi: 10.1104/pp.15.00994
- Morita, H., Yamashita, M., Shi, S. P., Wakimoto, T., Kondo, S., Kato, R., et al. (2011). Synthesis of unnatural alkaloid scaffolds by exploiting plant polyketide synthase. *Proc. Natl. Acad. Sci. U.S.A.* 108, 13504–13509. doi: 10.1073/pnas.1107782108
- Morita, Y., Takagi, K., Fukuchi-Mizutani, M., Ishiguro, K., Tanaka, Y., Nitasaka, E., et al. (2014). A chalcone isomerase-like protein enhances flavonoid production and flower pigmentation. *Plant J.* 78, 294–304. doi: 10.1111/tjp.12469
- Moummou, H., Kallberg, Y., Tonfack, L. B., Persson, B., and van der Rest, B. (2012). The plant short-chain dehydrogenase (sdr) superfamily: genome-wide inventory and diversification patterns. *BMC Plant Biol.* 12:219. doi: 10.1186/1471-2229-12-219
- Nelson, D., and Werck-Reichhart, D. (2011). A p450-centric view of plant evolution. *Plant J.* 66, 194–211. doi: 10.1111/j.1365-3113X.2011.04529.x
- Nelson, D. R. (2006). Plant cytochrome p450s from moss to poplar. *Phytochem. Rev.* 5, 193–204. doi: 10.1007/s11101-006-9015-3
- Ngaki, M. N., Louie, G. V., Philippe, R. N., Manning, G., Pojer, F., Bowman, M. E., et al. (2012). Evolution of the chalcone-isomerase fold from fatty-acid binding to stereospecific catalysis. *Nature* 485, 530–533. doi: 10.1038/nature11009
- Owens, D. K., Alerding, A. B., Crosby, K. C., Bandara, A. B., Westwood, J. H., and Winkel, B. S. (2008a). Functional analysis of a predicted flavonol synthase gene family in *Arabidopsis*. *Plant Physiol.* 147, 1046–1061. doi: 10.1104/pp.108.117457
- Owens, D. K., Crosby, K. C., Runac, J., Howard, B. A., and Winkel, B. S. (2008b). Biochemical and genetic characterization of *Arabidopsis* flavanone 3beta-hydroxylase. *Plant Physiol. Biochem.* 46, 833–843. doi: 10.1016/j.plaphy.2008.06.004
- Pichersky, E., and Gang, D. R. (2000). Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. *Trends Plant Sci.* 5, 439–445. doi: 10.1016/S1360-1385(00)01741-6
- Preuss, A., Stracke, R., Weisshaar, B., Hillebrecht, A., Matern, U., and Martens, S. (2009). *Arabidopsis thaliana* expresses a second functional flavonol synthase. *FEBS Lett.* 583, 1981–1986. doi: 10.1016/j.febslet.2009.05.006
- Rauscher, M. D. (2006). “The evolution of flavonoids and their genes,” in *The Science of Flavonoids*, ed E. Grotenwald (New York, NY: Springer), 175–211.
- Rauscher, M. D., Miller, R. E., and Tiffin, P. (1999). Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. *Mol. Biol. Evol.* 16, 266–274. doi: 10.1093/oxfordjournals.molbev.a026108
- Rauter, A. P., Ennis, M., Hellwich, K. H., Herold, B. J., Horton, D., Moss, G. P., et al. (2018). Nomenclature of flavonoids (IUPAC recommendations 2017). *Pure Appl Chem.* 90, 1429–1486. doi: 10.1515/pac-2013-0919
- Saito, K., Yonekura-Sakakibara, K., Nakabayashi, R., Higashi, Y., Yamazaki, M., Tohge, T., et al. (2013). The flavonoid biosynthetic pathway in *Arabidopsis*: structural and genetic diversity. *Plant Physiol. Biochem.* 72, 21–34. doi: 10.1016/j.plaphy.2013.02.001
- Schuz, R., Heller, W., and Hahlbrock, K. (1983). Substrate specificity of chalcone synthase from petroselinum hortense. Formation of phloroglucinol derivatives from aliphatic substrates. *J. Biol. Chem.* 258, 6730–6734.
- Seitz, C., Eder, C., Deiml, B., Kellner, S., Martens, S., and Forkmann, G. (2006). Cloning, functional identification and sequence analysis of flavonoid 3'-hydroxylase and flavonoid 3',5'-hydroxylase cDNAs reveals independent evolution of flavonoid 3',5'-hydroxylase in the *Asteraceae* family. *Plant Mol. Biol.* 61, 365–381. doi: 10.1007/s11103-006-0012-0
- Shimada, N., Aoki, T., Sato, S., Nakamura, Y., Tabata, S., and Ayabe, S. (2003). A cluster of genes encodes the two types of chalcone isomerase involved in the biosynthesis of general flavonoids and legume-specific 5-deoxy(iso)flavonoids in *Lotus japonicus*. *Plant Physiol.* 131, 941–951. doi: 10.1104/pp.004820
- Shimizu, Y., Ogata, H., and Goto, S. (2017). Type iii polyketide synthases: functional classification and phylogenomics. *ChemBioChem.* 18, 50–65. doi: 10.1002/cbic.201600522
- Shirley, B. W. (1996). Flavonoid biosynthesis: 'new' functions for an 'old' pathway. *Trends Plant Sci.* 1, 377–382. doi: 10.1016/1360-1385(96)10040-6
- Stafford, H. A. (1991). Flavonoid evolution: an enzymic approach. *Plant Physiol.* 96, 680–685. doi: 10.1104/pp.96.3.680
- St-Pierre, B., and De Luca, V. (2000). “Evolution of acyltransferase genes: origin and diversification of the bahl superfamily of acyltransferases involved in secondary metabolism,” in *Recent Advances in Phytochemistry*, eds J. T. Romeo, R. Ibrahim, L. Varin, and V. De Luca (Amsterdam: Elsevier BV), 285–315.
- Subramanian, S., Stacey, G., and Yu, O. (2007). Distinct, crucial roles of flavonoids during legume nodulation. *Trends Plant Sci.* 12, 282–285. doi: 10.1016/j.tplants.2007.06.006
- Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., et al. (2007). Proposed minimum reporting standards for chemical analysis. *Metabolomics* 3, 211–221. doi: 10.1007/s11306-007-0082-2
- Tohge, T., Watanabe, M., Hoefgen, R., and Fernie, A. R. (2013). The evolution of phenylpropanoid metabolism in the green lineage. *Crit. Rev. Biochem. Mol. Biol.* 48, 123–152. doi: 10.3109/10409238.2012.758083
- Tonfack, L. B., Moummou, H., Latché A., Youmbi, E., Benichou, M., Pech, J.-C., et al. (2011). The plant sdr superfamily: involvement in primary and secondary metabolism. *Curr. Topics Plant Biol.* 12, 41–53. Available online at: <https://prodnra.inra.fr/record/438703>
- Tsugawa, H., Kind, T., Nakabayashi, R., Yukihiro, D., Tanaka, W., Cajka, T., et al. (2016). Hydrogen rearrangement rules: computational ms/ms fragmentation and structure elucidation using ms-finder software. *Anal. Chem.* 88, 7946–7958. doi: 10.1021/acs.analchem.6b00770
- Tsugawa, H., Nakabayashi, R., Mori, T., Yamada, Y., Takahashi, M., Rai, A., et al. (2019). A cheminformatics approach to characterize metabolomes in stable-isotope-labeled organisms. *Nat. Methods.* 16, 295–298. doi: 10.1038/s41592-019-0358-2
- Tuominen, L. K., Johnson, V. E., and Tsai, C. J. (2011). Differential phylogenetic expansions in bahl acyltransferases across five angiosperm taxa and evidence of divergent expression among populus paralogues. *BMC Genomics* 12:236. doi: 10.1186/1471-2164-12-236
- van den Born, E., Omelchenko, M. V., Bekkelund, A., Leihne, V., Koonin, E. V., Dolja, V. V., et al. (2008). Viral alkB proteins repair rna damage by oxidative demethylation. *Nucleic Acids Res.* 36, 5451–5461. doi: 10.1093/nar/gkn519
- Wang, P., Zhang, L., Jiang, X., Dai, X., Xu, L., Li, T., et al. (2018). Evolutionary and functional characterization of leucoanthocyanidin reductases from *Camellia sinensis*. *Planta* 247, 139–154. doi: 10.1007/s00425-017-2771-z
- Wang, X., He, X., Lin, J., Shao, H., Chang, Z., and Dixon, R. A. (2006). Crystal structure of isoflavone reductase from alfalfa (*Medicago sativa* L.). *J. Mol. Biol.* 358, 1341–1352. doi: 10.1016/j.jmb.2006.03.022
- Wang, Y., Tan, X., and Paterson, A. H. (2013). Different patterns of gene structure divergence following gene duplication in *Arabidopsis*. *BMC Genomics* 14:652. doi: 10.1186/1471-2164-14-652
- Wanibuchi, K., Zhang, P., Abe, T., Morita, H., Kohno, T., Chen, G. S., et al. (2007). An acridone-producing novel multifunctional type iii polyketide synthase from *Hypericium serrata*. *FEBS J.* 274, 1073–1082. doi: 10.1111/j.1742-4658.2007.05656.x
- Weng, J. K., Li, X., Stout, J., and Chapple, C. (2008). Independent origins of syringyl lignin in vascular plants. *Proc. Natl. Acad. Sci. U.S.A.* 105, 7887–7892. doi: 10.1073/pnas.0801696105
- Weng, J. K., and Noel, J. P. (2012a). The remarkable pliability and promiscuity of specialized metabolism. *Cold Spring Harb. Symp. Quant. Biol.* 77, 309–320. doi: 10.1101/sqb.2012.77.014787
- Weng, J. K., and Noel, J. P. (2012b). Structure-function analyses of plant type iii polyketide synthases. *Methods Enzymol.* 515, 317–335. doi: 10.1016/B978-0-12-394290-6.00014-8
- Weng, J. K., and Noel, J. P. (2013). Chemodiversity in *Selaginella*: a reference system for parallel and convergent metabolic evolution in terrestrial plants. *Front Plant Sci.* 4:119. doi: 10.3389/fpls.2013.00119
- Weng, J. K., Ye, M., Li, B., and Noel, J. P. (2016). Co-evolution of hormone metabolism and signaling networks expands plant adaptive plasticity. *Cell* 166, 881–893. doi: 10.1016/j.cell.2016.06.027

- Williams, C. A., and Grayer, R. J. (2004). Anthocyanins and other flavonoids. *Nat. Product Rep.* 21, 539–573. doi: 10.1039/b311404j
- Winkel-Shirley, B. (2001). Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol.* 126, 485–493. doi: 10.1104/pp.126.2.485
- Wolf, L., Rizzini, L., Stracke, R., Ulm, R., and Rensing, S. A. (2010). The molecular and physiological responses of *Physcomitrella patens* to ultraviolet-b radiation. *Plant Physiol.* 153, 1123–1134. doi: 10.1104/pp.110.154658
- Wu, L. F., Meng, S., and Tang, G. L. (2016). Ferrous iron and alpha-ketoglutarate-dependent dioxygenases in the biosynthesis of microbial natural products. *Biochim. Biophys. Acta.* 1864, 453–470. doi: 10.1016/j.bbapap.2016.01.012
- Yazaki, K., Sasaki, K., and Tsurumaru, Y. (2009). Prenylation of aromatic compounds, a key diversification of plant secondary metabolites. *Phytochemistry* 70, 1739–1745. doi: 10.1016/j.phytochem.2009.08.023
- Yonekura-Sakakibara, K., and Hanada, K. (2011). An evolutionary view of functional diversity in family 1 glycosyltransferases. *Plant J.* 66, 182–193. doi: 10.1111/j.1365-313X.2011.04493.x
- Yu, C. K. Y., Springob, K., Schmidt, J. R., Nicholson, R. L., Chu, I. K., Yip, W. K., et al. (2005). A stilbene synthase gene (*sbsts1*) is involved in host and nonhost defense responses in sorghum. *Plant Physiol.* 138, 393–401. doi: 10.1104/pp.105.059337
- Yu, H. N., Liu, X. Y., Gao, S., Sun, B., Zheng, H. B., Ji, M., et al. (2018). Structural and biochemical characterization of the plant type iii polyketide synthases of the liverwort *Marchantia paleacea*. *Plant Physiol. Biochem.* 125, 95–105. doi: 10.1016/j.plaphy.2018.01.030
- Yu, X. H., Gou, J. Y., and Liu, C. J. (2009). Bahd superfamily of acyl-coa dependent acyltransferases in populus and arabidopsis: bioinformatics and gene expression. *Plant Mol. Biol.* 70, 421–442. doi: 10.1007/s11103-009-9482-1
- Yuan, T., Fujioka, S., Takatsuto, S., Matsumoto, S., Gou, X., He, K., et al. (2007). Ben1, a gene encoding a dihydroflavonol 4-reductase (*dfr*)-like protein, regulates the levels of brassinosteroids in *Arabidopsis thaliana*. *Plant J.* 51, 220–233. doi: 10.1111/j.1365-313X.2007.03129.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Yonekura-Sakakibara, Higashi and Nakabayashi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Atypical Myrosinase as a Mediator of Glucosinolate Functions in Plants

Ryosuke Sugiyama and Masami Y. Hirai*

RIKEN Center for Sustainable Resource Science, Yokohama, Japan

OPEN ACCESS

Edited by:

Jens Rohloff,
Norwegian University of Science
and Technology, Norway

Reviewed by:

Ute Wittstock,
Technische Universität Braunschweig,
Germany
Verena Jeschke,
University of Copenhagen, Denmark

*Correspondence:

Masami Y. Hirai
masami.hirai@riken.jp

Specialty section:

This article was submitted to
Plant Metabolism
and Chemodiversity,
a section of the journal
Frontiers in Plant Science

Received: 02 May 2019

Accepted: 18 July 2019

Published: 06 August 2019

Citation:

Sugiyama R and Hirai MY (2019)
Atypical Myrosinase as a Mediator
of Glucosinolate Functions in Plants.
Front. Plant Sci. 10:1008.
doi: 10.3389/fpls.2019.01008

Glucosinolates (GLSs) are a well-known class of specialized plant metabolites, distributed mostly in the order Brassicales. A vast research field in basic and applied sciences has grown up around GLSs owing to their presence in important agricultural crops and the model plant *Arabidopsis thaliana*, and their broad range of bioactivities beneficial to human health. The major purpose of GLSs in plants has been considered their function as a chemical defense against predators. GLSs are physically separated from a specialized class of beta-thioglucosidases called myrosinases, at the tissue level or at the single-cell level. They are brought together as a consequence of tissue damage, primarily triggered by herbivores, and their interaction results in the release of toxic volatile chemicals including isothiocyanates. In addition, recent studies have suggested that plants may adopt other strategies independent of tissue disruption for initiating GLS breakdown to cope with certain biotic/abiotic stresses. This hypothesis has been further supported by the discovery of an atypical class of GLS-hydrolyzing enzymes possessing features that are distinct from those of the classical myrosinases. Nevertheless, there is only little information on the physiological importance of atypical myrosinases. In this review, we focus on the broad diversity of the beta-glucosidase subclasses containing known atypical myrosinases in *A. thaliana* to discuss the hypothesis that numerous members of these subclasses can hydrolyze GLSs to regulate their diverse functions in plants. Also, the increasingly broadening functional repertoires of known atypical/classical myrosinases are described with reference to recent findings. Assessment of independent insights gained from *A. thaliana* with respect to (1) the phenotype of mutants lacking genes in the GLS metabolic/breakdown pathways, (2) fluctuation in GLS contents/metabolism under specific conditions, and (3) the response of plants to exogenous GLSs or their hydrolytic products, will enable us to reconsider the physiological importance of GLS breakdown in particular situations, which is likely to be regulated by specific beta-glucosidases.

Keywords: glucosinolate, myrosinase, beta-glucosidase, metabolism, stress response

INTRODUCTION

Over the years, a number of bioactive metabolites have been identified in plants, many of which are utilized as beneficial sources of pharmaceuticals and/or research tools. Glucosinolate (GLS), a class of sulfur-rich natural products mainly produced by the family Brassicaceae, is among the most studied plant metabolites owing to its potential health-related benefits and availability in the model plant *Arabidopsis thaliana* (Halkier and Gershenzon, 2006; Agerbirk and Olsen, 2012). GLSs impart

specific pungency and flavors to Brassicaceae vegetables such as mustard and cabbage (Fahey et al., 2001; Halkier, 2016; Possenti et al., 2017). Moreover, a few GLS compounds such as glucoraphanin (4-methylsulfinyl-*n*-butyl glucosinolate) are known to produce health-promoting chemicals with diverse bioactivities (Traka, 2016; Banerjee and Paruthy, 2017). Very recently, the enormous body of research activities on GLS has been compiled as two books entitled *Glucosinolates* with few overlaps (Kopriva, 2016; Mérillon and Ramawat, 2017).

In recent years, chemical ecology, which focuses on gaining an understanding of the physiological functions of specialized metabolites (previously referred to as secondary metabolites) in organisms, has also become a burgeoning scientific field in natural product research. In this context, GLSs have traditionally been considered as defense chemicals deployed against predators. GLSs generally accumulate in specific cells (S-cells), separated from cells containing their hydrolytic enzymes (beta-thioglucosidases called myrosinases) (Halkier, 2016; Wittstock et al., 2016a). In addition, it has been suggested that GLSs and myrosinases could co-exist even within single cells, probably being compartmentalized in different organelles (Koroleva and Cramer, 2011). They are mixed upon tissue damage to release toxic volatiles such as isothiocyanates (ITCs) (Figure 1). Specifier proteins and side chain structures play an important role in converting the unstable aglycon to various end products with different bioactivities (Lambrix et al., 2001; Hanschen et al., 2014; Wittstock et al., 2016a,b; Eisenschmidt-Bönn et al., 2019). For instance, simple nitriles are generated in the presence of nitrile specifier proteins, and ITCs possessing a hydroxyl group at position 2 can be further cyclized. These compartmentalizations, which enable plants to safely control harmful chemicals, is referred to as the GLS-myrosinase system or “mustard oil bomb” and has long fascinated many plant scientists (Lüthy and Matile, 1984).

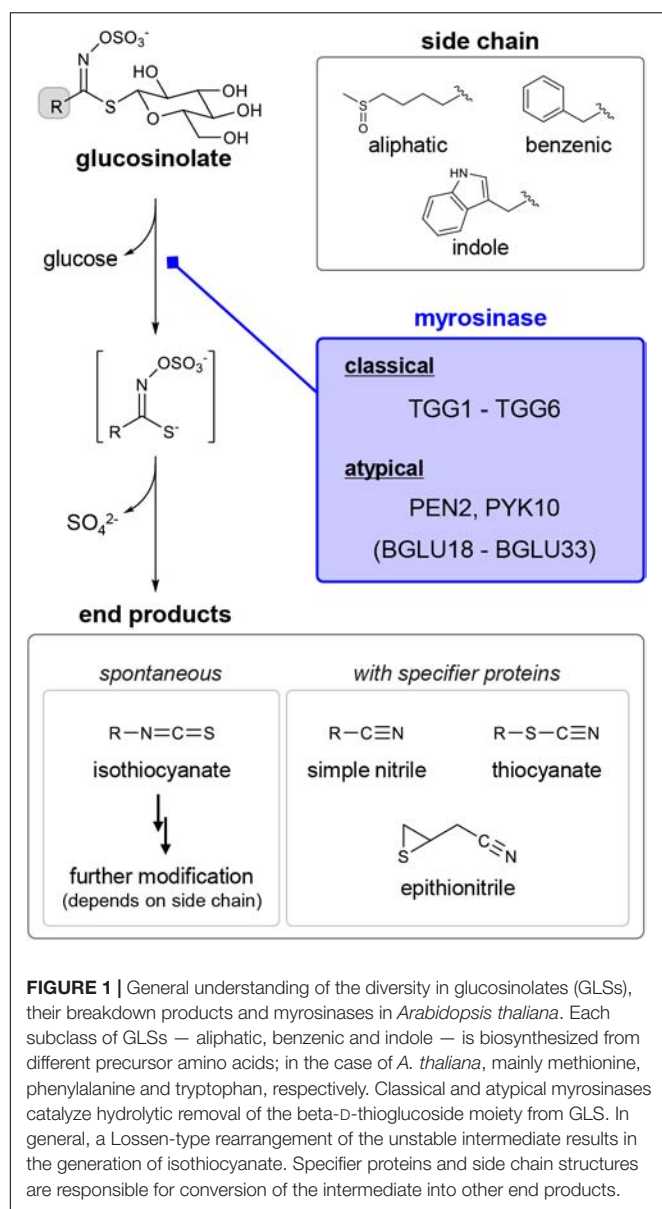
Knowledge on the chemical diversity embedded in the GLS metabolism and on the molecular mechanisms underlying the GLS-myrosinase system is frequently updated. For example, energetical investigations have been made of GLS biosynthetic genes (Halkier, 2016; Barco and Clay, 2019), end products directed by specifier proteins (Wittstock et al., 2016a), tissue localization via GLS transporters (Jørgensen et al., 2015; Halkier, 2016), differences in GLS contents among species and accessions (Kliebenstein and Cacho, 2016), and proteins that interact with myrosinases to regulate their activity and stability (Bhat and Vyas, 2019; Chen et al., 2019). However, most of the insights that have been gained regarding the molecular basis and physiological importance of GLS breakdown are based on the intercellular and tissue damage-dependent GLS-myrosinase system. On the other hand, there are several reports on fluctuations of endogenous GLS levels even in non-disrupted tissues, caused by pathogen attack or abiotic stress (Martinez-Ballesta et al., 2013; Variyar et al., 2014; Burow, 2016; Pastorczyk and Bednarek, 2016). These observations suggest the existence of different system(s) that regulate GLS breakdown independent of tissue disruption, to cope with such

environmental stresses. In connection with the subcellular GLS-myrosinase compartmentalization, additional functions of GLSs not limited to their role as the defense chemicals against herbivores have been recognized (Bednarek, 2012b; Katz et al., 2015; Francisco et al., 2016b; Burow and Halkier, 2017). In fact, many of the latest studies have demonstrated broad physiological functions of GLSs (Francisco et al., 2016a; Katz and Chamovitz, 2017; Malinovsky et al., 2017; Nintemann et al., 2018; Urbancsok et al., 2018a,b).

In order to gain a more comprehensive understanding of the possible multi-functionality of GLSs *in planta*, we should take a global view of the profound diversification embedded in GLSs (Figure 1). Nearly 150 GLS compounds have been identified to date, and there are at least 36 GLSs with different side-chain structures in *A. thaliana* (Brown et al., 2003; Agerbirk et al., 2018). Three subclasses of GLSs — aliphatic, benzenic and indole GLSs — are biosynthesized from different precursor amino acids with independent regulatory systems by MYB and MYC transcription factors (Figure 1) (Frerigmann, 2016). Thus, it is conceivable that each GLS class could participate in distinct biological processes, as indole GLSs are known to play an essential role in plant immunity via coordination with the metabolism of other phytoalexins (Bednarek et al., 2009; Pedras et al., 2011; Klein and Sattely, 2015, 2017; Frerigmann et al., 2016). In addition, different bioactivities of ITCs, dependent on their side-chain structures, are well recognized despite their non-specific nucleophilicity, which may contribute to fitness performance of plants (Burow et al., 2010; Andersson et al., 2015; Urbancsok et al., 2017). Moreover, the production of various end products even from a single GLS species, directed by the specifier proteins, is likely to expand the endogenous biological targets of GLSs in different signaling pathways (Lambrix et al., 2001; Wittstock et al., 2016a,b). In contrast, the genetic and biochemical diversity in myrosinases is less understood, even though a new class of beta-glucosidases capable of hydrolyzing GLSs has been identified. Compared with the well-documented class of myrosinases that are widely found in the order Brassicales, these so-called atypical myrosinases possess unique features with respect to both amino acid sequences and enzymatic profiles. In this review, we therefore focus on the considerable diversity of beta-glucosidases in *A. thaliana* as a model to discuss their possible contribution to the broad utility of GLSs in plants, even at the subcellular level.

CLASSICAL VERSUS ATYPICAL MYROSINASES

In *A. thaliana*, glucosyl hydrolase family I is composed of 47 *BETA-GLUCOSIDASE* (*BGLU*) genes and a *BGLU*-like gene, *AFR2* (Xu et al., 2004). According to the crystal structure of a thioglucosidase from *Sinapis alba*, a few amino acid residues are conserved in myrosinases among a wide range of GLS-producing plants (Burmeister et al., 1997). For example, Gln and Glu residues within the catalytic site are considered to be essential for cleavage of the



thioglucoside moiety. Thus, six genes (*BGLU34–BGLU39*) named *THIOGLUCOSIDE GLUCOHYDROLASE* (*TGG*) had previously been considered as the only class encoding myrosinases in *A. thaliana*. Myrosinases possessing these amino acid signatures have been found in a wide range of GLS-producing plants (Rask et al., 2000).

In 2009, however, it was revealed that PENETRATION 2 (PEN2)/*BGLU26* is capable of hydrolyzing indole GLSs and that generation of the putative degradation products is critical for the plant immune response (Bednarek et al., 2009; Clay et al., 2009). Although the key Gln residue is replaced by Glu in PEN2, the recombinant PEN2 protein clearly showed myrosinase activity against indol-3-ylmethyl glucosinolate (I3G) and its 4-methoxy analog (Bednarek et al., 2009). Moreover, Nakano et al. (2014) demonstrated that PYK10/*BGLU23* is a

major component of the endoplasmic reticulum (ER) body — an organelle found primarily in the family Brassicaceae — and also has functions as a myrosinase against I3G (Nakano et al., 2017). PYK10 also has Glu instead of the key Gln residue found in TGGs. Based on these findings, PEN2 and PYK10 were newly categorized as atypical or EE-type myrosinases, in contrast to TGG1–TGG6, which are referred to as classical or QE-type myrosinases (“EE” and “QE” represent their conserved amino acid residues).

It should be noted that the two Glu residues identified in PEN2 and PYK10 are conserved among the 16 genes named *BGLU18–BGLU33* in *A. thaliana* (Table 1). A phylogenomic analysis of BGLUs from more than 50 plant species revealed that the monophyletic clade composed of these 16 BGLUs is specific for the order Brassicales (Nakano et al., 2017). These BGLU members lack the Gln and other amino acid signatures conserved in the classical myrosinases, whereas additional basic residues oriented to the deduced substrate-binding pocket occur only in this subclass. Detailed amino acid signatures in *A. thaliana* BGLUs are shown in Nakano et al. (2017), especially in their Figure 5 and Supplementary Figure S11. Considering the distinct amino acid signatures conserved in each BGLU subclass and their frequent emergence across plant species, it is suggested that the QE and EE myrosinases have arisen independently during evolution (Nakano et al., 2017). Not limited to PEN2 and PYK10, interestingly, transcriptional changes and mutations in some of these BGLUs have suggested their relevance in response to specific stresses (Figure 2, Tables 1, 2, and Supplementary Tables S2, S3). It is also to be noted that GLS metabolism is affected independent of tissue disruption under those conditions (Table 2). Therefore, other members of this BGLU subclass may have myrosinase activities and regulate different machineries for GLS turnover that are perhaps more specialized in substrate selectivity, tissue localization and developmental stage, rather than the broad-scale chemical defense against herbivores deployed by classical myrosinases.

In this review, we aim to discuss the hypothesis that a wide range of these Brassicales-specific BGLUs can function as myrosinases to regulate the multiple functions of GLSs in planta. Although current insights into this BGLU subclass are highly limited, essentially three types of previous studies could be useful for considering this hypothesis: analyses of (1) the phenotype of mutants lacking genes responsible for the GLS metabolic/breakdown pathway; (2) the changes in GLS levels/metabolism under specific conditions; and (3) the response of plants treated with GLSs or their hydrolytic products (Table 2). Here, mainly based on studies in *A. thaliana*, we consider phenotypic information related to *BGLU18–BGLU33*, fluctuations in GLS metabolism, and the effects of exogenous GLS breakdown products on plants to review the physiological importance of GLS breakdown in particular situations. In addition to a specific focus on atypical myrosinases, more diverse functions of classical TGGs are described with reference to the most recent insights. Finally, we suggest the types of experiment that are effective in gaining

TABLE 1 | Current insights on BGLU18–BGLU39, putative EE-type myrosinases in *Arabidopsis thaliana*.

Class ^a	Gene name	AGI ID	Substrates ^b	Subcellular localization	Phenotypes of knock-down mutants	Notes
III	BGLU18/BG1	At1g52400	ABA-GE ¹ (4MI3G) ²	ER body ³	Susceptible to <i>A. vulgare</i> attacks (<i>pyk10 bglu18</i>) ² susceptible to drought ¹ Tolerant to salt stress ⁵	A major component of inducible ER bodies ⁴ Up-regulation by high NaCl ⁵
	BGLU19	At3g21370				
	BGLU20/ATA27	At1g75940				
	BGLU21	At1g66270	Coumarin glucosides ⁶	ER body ³		A component of root ER bodies ⁷ cannot hydrolyze sinigrin ⁶
	BGLU22	At1g66280	Coumarin glucosides ⁶	ER body ³		A component of root ER bodies ⁷ cannot hydrolyze sinigrin ⁶
	BGLU23/PYK10	At3g09260	Coumarin glucosides ⁶ I3G , ⁸ (4MI3G) ²	ER body ³	Susceptible to <i>A. vulgare</i> attacks (<i>pyk10 bglu18</i>) ²	A major component of root and leaf ER bodies ⁷ cannot hydrolyze sinigrin ⁶ Very low signals in all tissues ⁸
	BGLU24	At5g28510				
IV	BGLU25/GLUC	At3g03640				
	BGLU26/PEN2	At2g44490	indole GLSs ⁹ 4-MUG ⁹	Peroxisome ¹⁰	Susceptible to pathogen attacks ¹⁰	
V	BGLU27	At3g60120				Very low signals in all tissues ⁸
	BGLU28	At2g44460				Up-regulation by sulfur depletion ¹¹
	BGLU29	At2g44470				
	BGLU30/ SRG2/DIN2	At3g60140				Up-regulation by sulfur depletion, ¹¹ extended darkness ¹² or senescence ¹³
	BGLU31	At5g24540				Very low signals in all tissues ⁸
VI	BGLU32	At5g24550				Very low signals in all tissues ⁸
	BGLU33/BG2	At2g32860	ABA-GE ¹⁴	Vacuole ¹⁴	Susceptible to salinity ¹⁴	

^aXu et al. (2004). ^bGlucosinolates are bolded. ABE-GE, abscisic acid glucosyl ester; 4MI3G, 4-methoxyindol-3-ylmethyl glucosinolate; I3G, indol-3-ylmethyl glucosinolate; 4-MUG, 4-methylumbelliferyl-beta-D-O-glucoside. ¹Lee et al. (2006), ²Nakazaki et al. (2019), ³Yamada et al. (2011), ⁴Ogasawara et al. (2009), ⁵Cao et al. (2017), ⁶Ahn et al. (2010), ⁷Matsushima et al. (2003), Nagano et al. (2008), ⁸Schmid et al. (2005), Nakabayashi et al. (2005), ⁹Bednarek et al. (2009), ¹⁰Lipka et al. (2005), ¹¹Hirai et al. (2003, 2004), Nikiforova et al. (2003), Maruyama-Nakashita et al. (2003), ¹²Fujiki et al. (2001), ¹³Lee J. et al. (2007), ¹⁴Xu et al. (2012).

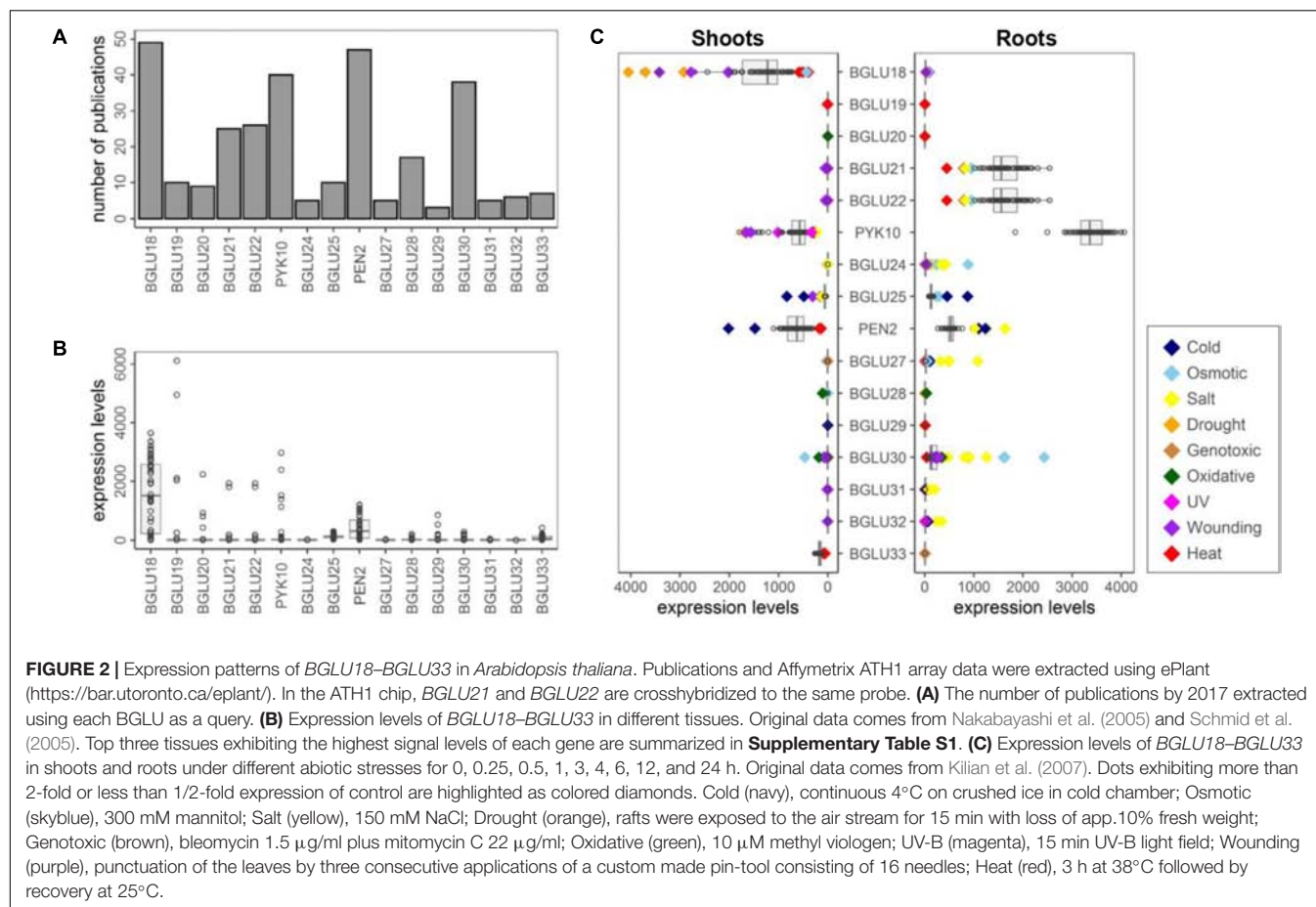
a more complete understanding of the multi-functionality of GLSs in plants.

A PLANT IMMUNE PATHWAY REGULATED BY PEN2, THE FIRST ATYPICAL MYROSINASE

Here, we describe the general understanding of the PEN2 pathway in brief, as the relevance of PEN2 and indole GLSs in plant immunity has been well documented (Bednarek, 2012a; Johansson et al., 2014; Frerigmann et al., 2016; Pastorczyk and Bednarek, 2016; Xu et al., 2016). PEN2 was first identified as a component of the pre-invasive resistance in *A. thaliana* against the powdery mildew fungi, *Blumeria graminis* f. sp. *hordei* and *Erysiphe pisi* (Lipka et al., 2005; Bednarek et al., 2009). PEN2 is also responsible for the callose deposit in *A. thaliana* seedlings induced by fungal pathogens or flg22, a bacterial flagellin-derived peptide (Clay et al., 2009), even though it is suggested that penetration resistance and callose deposition are not directly linked (Lipka et al., 2005; Maeda et al., 2009;

Hiruma et al., 2010). These stimuli induce a decrease of indole GLS levels mediated by PEN2 and the accumulation of plausible end products including indol-3-ylmethylamine (I3A), raphanusamic acid (RA), and 4-O-beta-D-glucosyl-indol-3-yl formamide (4OGlcI3F) (Bednarek et al., 2009, 2011; Lu et al., 2015). Notably, hydroxylation of I3G at position 4 mediated by CYP81F2 is a critical step for the PEN-dependent immune response. This is an interesting example of a GLS with a particular side chain exhibiting a specialized biological function. On the other hand, the unstable indol-3-ylmethyl ITC can serve as an intermediate for the biosynthesis of several phytoalexins (Bednarek et al., 2009; Klein and Sattely, 2017). Thus, the molecular mechanisms and actual bioactive metabolite(s) relevant in this pathway remain to be further investigated.

In a recent study, a co-expression analysis for genes involved in the PEN2 immune system identified GLUTATHIONE S-TRANSFERASE CLASS-TAU MEMBER 13 (GSTU13) as a critical component of that pathway (Piślewska-Bednarek et al., 2018). The *gstu13* mutants were more susceptible to several pathogens and were impaired in



callose deposition induced by the bacterial flg22 epitope. Furthermore, the formation of pathogen-triggered specific metabolites such as indol-3-ylmethyl amine was broadly repressed in these mutants. Therefore, the conjugation of ITCs with glutathione catalyzed by GSTU13 was revealed to be strictly essential for the activation of the indole GLS-related immune system.

PYK10, BGLU18 AND ER-RETAINED BGLUs COULD MAINTAIN AN INTRACELLULAR GLS-MYROSINASE SYSTEM

In the GLS-myrosinase system, physical separation of glucose-conjugated precursor compounds from their hydrolases is an efficient means to control the bioactivity of these metabolites, and this strategy may function even at the subcellular level. The ER body, a rod-shaped organelle continuous with the ER, is considered to provide such an intracellular compartment; a few BGLUs including PYK10/BGLU23 are significantly enriched in these structures in *A. thaliana* (Matsushima et al., 2003; Ogasawara et al., 2009). Among the Brassicales-specific BGLUs, BGLU18-BGLU25 commonly have

ER-retention signals in their signal peptides and C-terminal regions (Nakano et al., 2014). Indeed, ER bodies that are constitutively present in roots accumulate large amounts of PYK10 and the closest homologs BGLU21 and BGLU22, whereas BGLU18 is the major component of another class of ER body that is induced by wounding or methyl jasmonate treatment (Matsushima et al., 2002, 2003; Nagano et al., 2008; Ogasawara et al., 2009). More detailed information on the physiology and molecular network in ER bodies is available in specific reviews (Yamada et al., 2011; Nakano et al., 2014; Shirakawa and Hara-Nishimura, 2018).

Notably, ER bodies have been observed in only a few families of the order Brassicales, namely, Brassicaceae, Capparaceae, and Cleomaceae. Therefore, substrate(s) of the BGLUs could also be restricted to a narrow range of phylogenetic clades. Although PYK10, BGLU21, and BGLU22 neither have the amino acid signatures conserved in classical myrosinases nor display myrosinase activities toward aliphatic allyl GLS (sinigrin) *in vitro* (Ahn et al., 2010), previous studies on PEN2 led Nakano and co-authors to hypothesize that PYK10 can hydrolyze indole GLSs. As expected, root protein extracts of the *pyk10* and several ER-body mutants in *A. thaliana*, as well as the recombinant PYK10 protein, showed myrosinase activity against I3G (Nakano et al., 2017). Interestingly, co-expression analysis revealed that PYK10 is more closely related

TABLE 2 | Relevance of glucosinolates (GLSs) and the Brassicales-specific beta-glucosidases (BGLUs) in *Arabidopsis thaliana* under abiotic stress.

Stress	GLS levels	Related BGLUs	Involvement in GLS metabolism	Effects of ITC treatment ^a
Drought	Decrease ¹	BGLU18 ² TGG1, TGG2 ³	Guard cells accumulate a large amount of TGG1 and TGG2. ³	AITC induces stomatal closure. ⁴
Salinity	Increase ⁵	BGLU19, ⁶ BGLU33 ⁷	<i>myb28 myb29</i> is more susceptible to salt stress. ⁵	–
Sulfur deficiency	Decrease ⁸	BGLU28, ⁹ BGLU30 ⁹	Growth of <i>gtr1 gtr2</i> is defected under low sulfur. ¹⁰	–
Light/Dark	Increase by light/ Decrease in the dark ¹¹	BGLU30 ¹² TGG1, TGG2 ¹³	GLS biosynthetic genes shows the diurnal rhythm. ¹¹	–
Temperature	–	Not determined	A low-GLS mutant (TU8) is more susceptible to heat stress. ¹⁴	AITC and PEITC enhances heat tolerance. ¹⁵

^aAITC, allyl isothiocyanate; PEITC, phenethyl isothiocyanate. ¹Ren et al. (2009), ²Lee et al. (2006), ³Zhao et al. (2008), ⁴Islam et al. (2009), ⁵Khokon et al. (2011), ⁶Martinez-Ballesta et al. (2015), ⁷Cao et al. (2017), ⁸Xu et al. (2012), ⁹Falk et al. (2007), ¹⁰Hirai et al. (2003, 2004), ¹¹Maruyama-Nakashita et al. (2003), ¹²Nikiforova et al. (2003), ¹³Nour-Eldin et al. (2012), ¹⁴Huseby et al. (2013), ¹⁵Fujiki et al. (2001), ¹⁶Brandt et al. (2018), ¹⁷Haughn et al. (1991), ¹⁸Ludwig-Müller et al. (2000), ¹⁹Hara et al. (2012).

to biosynthetic genes for indole GLSs than to those of coumarin glucosides, putative substrates predicted from *in vitro* assays (Ahn et al., 2010; Nakano et al., 2017). Therefore, the physiological function of PYK10 *in planta* is more likely to be associated with GLS metabolism. Furthermore, a suite of informatics analyses performed in Nakano et al. (2017) indicated that a broad class of BGLUs, not limited to the classical myrosinases, may have the potential to hydrolyze GLSs, as described above.

The myrosinase activity of PYK10 toward indole GLSs is further supported by a very recent study on a new class of ER bodies, referred to as leaf ER bodies (Nakazaki et al., 2019). Compared with the aforementioned known classes of ER bodies, leaf ER bodies occur constitutively in a few types of epidermal cells in rosette leaves and accumulate both PYK10 and BGLU18. The *pyk10 bglu18* mutant was shown to lack leaf ER bodies and became more susceptible to attack by the terrestrial isopod *Armadillidium vulgare* compared with the wild type plants (Nakazaki et al., 2019). In addition, the levels of most endogenous GLS species decrease rapidly in homogenates of the rosette leaves, mainly due to the activities of TGG1 and TGG2, whereas degradation of 4-methoxyindol-3-ylmethyl glucosinolate (4MI3G) was found to be selectively and significantly delayed in the *pyk10 bglu18* mutant (Nakazaki et al., 2019). 4MI3G is a key GLS species in the PEN2-dependent immune pathway (Bednarek et al., 2009; Clay et al., 2009). Given that damage to leaves of the *tgg1 tgg2* mutant caused by *A. vulgare* in a feeding assay was comparable to that in the wild type, it has been suggested that leaf ER bodies are involved in the production of the defensive chemicals from 4MI3G that protect *A. thaliana* leaves against herbivore attack. Since neither *pyk10* nor *bglu18* single mutants were examined in these experiments, it would be of interest to determine whether BGLU18 can also hydrolyze indole GLSs. In combination with previous findings, these findings suggested that ER bodies can provide a further class of GLS–myrosinase compartment at the subcellular level (i.e., GLSs retained in vacuoles and myrosinases retained in the ER bodies), that

plays an important role in plant defense against attacks of herbivores and pathogens.

DO BGLU18 AND BGLU33 HAVE OTHER SUBSTRATES IN ADDITION TO GLUCOSE-CONJUGATED ABSCISIC ACID?

In addition to its potential roles in the ER bodies, BGLU18 is known to participate in abscisic acid (ABA) metabolism. ABA, one of the most important phytohormones active during a plant's life cycle, is involved in a variety of biological processes, including the adaptation to environmental stresses (Sakata et al., 2014; Daszkowska-Golec, 2016). The cellular ABA level is partially regulated via a complex *de novo* biosynthetic pathway (Nambara and Marion-Poll, 2005). In addition, the discovery of ABA-glucosyltransferase, which generates an ABA glucosyl ester (ABA-GE), led us to hypothesize that release of ABA from the pool of inactive ABA analogs can potentially modulate ABA concentrations more dynamically (Xu et al., 2002). BG1/BGLU18 was reported as the first enzyme that can hydrolyze ABA-GE as part of the drought stress response (Lee et al., 2006). Subsequently, BG2/BGLU33 was identified as a further member of the ABA-GE hydrolases localized in vacuoles and was shown to play an important role under conditions of salt stress (Xu et al., 2012). Although the enzymatic potential of these BGLUs to hydrolyze ABA-GE and their relevance in ABA functions have been well investigated by independent groups (Merilo et al., 2015; Ondzighi-Assoume et al., 2016; Yamashita et al., 2016), it is questionable whether ABA-GE is the sole substrate of these BGLUs under physiological conditions, based on a consideration of the following observation.

In the first place, the response to drought and salinity is subject to complex regulation, not only by ABA but also by other small molecules including GLSs. Their contribution to the drought stress response has been discussed with regard to

stomatal closure in guard cells, which is the major response of plants under low water conditions. The signal cascade for stomatal movement is initiated by multiple inputs from plant hormones and several primary/specialized metabolites in response to environmental stresses, whereas they are finally integrated into a single output — the production of reactive oxygen species (ROS) and Ca^{2+} oscillation followed by protein phosphorylation (Murata et al., 2015). Hence, depletion of one signal molecule could be compensated by the function of another compound. Notably, application of exogenous allyl ITC or several GLS breakdown products has been shown to induce stomatal closure in *A. thaliana* leaves (Khokon et al., 2011; Hossain et al., 2013). Furthermore, TGG1 and TGG2 have been reported to be major components of guard cells and shown to be involved in stomatal movement (Zhao et al., 2008; Islam et al., 2009). The stomatal closure induced by allyl ITC and the subsequent ROS production does not require endogenous ABA, but is dependent on methyl jasmonate (Khokon et al., 2011), indicating that ABA and ITC function as independent inputs for stomatal movement. Some researchers have hypothesized that the accumulation of TGGs and GLSs in guard cells represents the evolutionary origin of the GLS–myrosinase system, because stomata can serve as the initial gateway to bacterial invasion (Shirakawa and Hara-Nishimura, 2018). However, whether breakdown of internal GLSs in guard cells occurs under drought conditions to induce stomatal closure and the relevance of BGLU18 in this process are still to be investigated. Involvement of GLSs in the salt stress response has also been suggested based on the findings of several studies that have examined the fluctuation of GLS contents in Brassicaceae plants and the response of *A. thaliana* mutants lacking aliphatic GLSs under salinity stress (López-Berenguer et al., 2008; Keling and Zhu, 2010; Martinez-Ballesta et al., 2015), even though the detailed mechanisms remain unclear.

Secondly, the EE-type myrosinases tend to have a dual function as S- and O-glucosidases. PYK10 has been reported to hydrolyze coumarin glucosides such as scopolin *in vitro* (Ahn et al., 2010), whereas co-expression analysis has indicated that indole GLSs are more likely to be the actual substrates *in planta* (Nakano et al., 2017). PEN2 is also known to have enzymatic potential to catalyze the deglycosylation of 4-methylumbelliferyl-beta-D-O-glucoside, albeit at a lower rate than for the hydrolysis of indole GLSs (Bednarek et al., 2009). Drought stress appears to promote the degradation of a wide range of GLS species not limited to indole GLSs (Ren et al., 2009). One possible reason is that GLSs are hydrolyzed after transportation to the compartments/cells containing classical myrosinases. Alternatively, as is the case for GLS degradation induced by sulfur depletion or prolonged darkness (see the following sections), it is also conceivable that EE-type myrosinases, which may include BGLU18 and BGLU33, could hydrolyze other GLS subclasses.

Thirdly, whereas ABA is one of the indispensable hormone compounds in the plant kingdom, the BGLU subclass containing BGLU18 and BGLU33 is distributed only in the order Brassicales (Xu et al., 2004; Nakano et al., 2017). Notably, BGLU18 is a major and essential component of ER bodies, an organelle observed in only a few families of the order Brassicales, as described above. One possibility is that BGLU18 and BGLU33 may have other

substrate(s) restricted to these small evolutionary clades, like PYK10. Another possible explanation is that there are several molecular systems to release ABA from the repository of inactive ABA derivatives, and these BGLUs may belong to one of those conserved only in the Brassicales. As a similar case, in some Brassicaceae plants, the metabolism of indole GLSs is closely related to that of auxin, another essential phytohormone (Malka and Cheng, 2017; Vik et al., 2018). Although similar machineries for dynamic regulation of ABA levels may exist in a wide range of plant families, such enzymes hydrolyzing inactive ABA derivatives have yet to be identified.

Taken together, the aforementioned findings indicate that we should not exclude the hypothesis that BGLU18 and BGLU33 can also function as myrosinases. If these BGLUs were proven to be dual-functional, the regulatory mechanism of these enzymatic activities *in planta* would inevitably attract greater attention.

GLS BREAKDOWN TO RECYCLE SULFUR MAY BE MEDIATED BY BGLU28 and BGLU30

This and the next section discuss the possibility that GLS itself could work as nutrient storage, not only as a precursor of toxic chemicals and signaling molecules. Brassicaceae plants containing GLSs need larger amounts of sulfur than other plants (Castro et al., 2003; Walker and Booth, 2003; Yang et al., 2006). Since GLSs have at least two sulfur atoms in each molecule and represent 10–30% of the total sulfur content in plant organs, these metabolites have been considered a potential source of sulfur for other metabolic processes (Falk et al., 2007). The effects of sulfur supply and depletion on GLS levels in plants have been well studied, including in agricultural practice (Falk et al., 2007; Schnug and Haneklaus, 2016). In *A. thaliana*, sulfur deficiency strongly induces down-regulation of GLS content as well as expression levels of GLS biosynthetic genes (Hirai et al., 2005; Zhang et al., 2011). To date, a few transcription factors, SULFUR LIMITATION 1 (SLIM1), SULFUR DEFICIENCY-INDUCED 1 (SDI1) and SDI2, have been reported to work as repressors of GLS biosynthesis under low sulfur conditions (Maruyama-Nakashita et al., 2006; Aarabi et al., 2016). In addition, developmental defects of GLS-less seeds as a result of mutations in GLS transporters (*gtr1 gtr2*) under sulfur deficiency further supported the potential role of GLSs as a sulfur reserve (Nour-Eldin et al., 2012). Nevertheless, most mechanisms underlying GLS turnover in that condition are still not known. It would be desirable to gain more direct evidence for the re-distribution of sulfur atoms from GLSs, e.g., by incorporating isotopes into primary sulfur metabolites from labeled GLSs.

Based on their observed up-regulation, BGLU28 and BGLU30 are suggested to be relevant in the hydrolysis of GLSs caused by sulfur depletion (Hirai et al., 2003, 2004; Maruyama-Nakashita et al., 2003; Nikiforova et al., 2003). BGLU28 and BGLU30 form a sister clade close to those containing PEN2 and PYK10 in the phylogenetic tree of *A. thaliana* BGLUs (Xu et al., 2004). Although their possible contribution to GLS turnover was suggested more than 15 years ago (Hirai and Saito, 2004),

there have been few reports on the physiological functions of these genes. To our knowledge, only two studies (Zhang et al., 2014; Jackson et al., 2015) have monitored *BGLU28* promoter activity, using the GUS reporter gene as a marker of low-sulfur response upon plant hormone treatment and mutations in *SULTR1;2*. In this context, it should be noted that sulfur deficiency primarily affects contents of aliphatic GLSs, whereas PYK10 and PEN2 have been reported to hydrolyze only indole GLSs. If *BGLU28* and *BGLU30* can indeed hydrolyze aliphatic GLSs, the broad chemodiversity of EE-type myrosinases would undoubtedly gain more recognition. Compared with the TGGs that exhibit low selectivity for GLS species (Zhou et al., 2012), EE-type myrosinases may have narrower substrate specificities and thus regulate only particular biotic/abiotic stress responses. Both the protein functions and physiological roles of these BGLUs need to be further investigated beyond the context of sulfur assimilation.

ACTIVATION OF GLS TURNOVER IN DARKNESS AND POSSIBLE CONTRIBUTION OF *BGLU30* TO RECOVER CARBOHYDRATES

Light is an essential energy source for the development and metabolism of higher plants, and the expression levels of a number of genes are known to be regulated by light; for example, the sulfur assimilation pathway is activated during the light period (Kocsy et al., 1997; Kopriva et al., 1999; Pruneda-Paz and Kay, 2010). Although the photo-regulation of sulfur assimilation has been investigated in different species under a variety of growth conditions (Buwalda et al., 1988; Passera et al., 1989; Lee E.-J. et al., 2007; Lee et al., 2011), there tends to be little consensus regarding its coordination with GLS biosynthesis/catabolism.

The regulation of GLS metabolism by light has been discussed from two aspects, namely, the substantial degradation of GLSs under conditions of prolonged darkness and the diurnal fluctuation of GLS contents. In *A. thaliana*, there is a marked reduction in the GLS content of leaves under conditions of extended darkness (Huseby et al., 2013; Brandt et al., 2018). Given that inhibition of photosynthesis is a strong stimulus inducing carbohydrate starvation and leaf senescence, GLS degradation may be a mechanism designed to cope with nutrient starvation via the mobilization of D-glucose units from those molecules. In this regard, *BGLU30*, also referred to as *DARK-INDUCIBLE 2 (DIN2)* or *SENESCENCE-RELATED GENE 2 (SRG2)*, is known to be significantly up-regulated in response to prolonged darkness, senescence, and sugar starvation (Fujiki et al., 2001; Lee J. et al., 2007). The key factor for the induction of *BGLU30* expression appears to be endogenous sugar levels rather than light conditions. A high abundance of *BGLU30* transcript was detected in detached leaves even under illumination in the presence of a photosynthesis inhibitor, 3-(3,4-dichlorophenyl)-1,1-dimethylurea, whereas it was barely detected under co-treatment with sucrose (Fujiki et al., 2001). Given that the *BGLU30* expression is also induced

by sulfur depletion, this enzyme may regulate a release of stored GLSs to overcome nutrient starvation under various conditions. Notably, the expression of *BGLU28* is induced neither by prolonged darkness nor by sugar starvation. In addition to their enzymatic functions, the difference in the regulatory systems between these closely related BGLUs is of particular interest.

Cooperating with the sulfur assimilation pathways, GLS levels have been shown to be higher during the day than at night in *A. thaliana* (Huseby et al., 2013). Simultaneously, the expression levels of GLS biosynthetic genes as well as the incorporation of inorganic ^{35}S into GLSs were found to be enhanced by light. Moreover, a further study revealed a diurnal increase in total myrosinase activity and the abundance of TGG1 and TGG2 proteins in *A. thaliana* seedlings (Brandt et al., 2018). These findings accordingly indicate that GLS metabolism is highly co-regulated with sulfur assimilation in response to the circadian rhythm, even though the physiological importance of this phenomenon remains unclear. Since the correlation between GLS contents and the expression levels of GLS biosynthetic genes was relatively low during the light period (Rosa, 1997; Klein et al., 2006; Schuster et al., 2006), not only *de novo* biosynthesis but also turnover could regulate the diurnal rhythm of endogenous GLSs. In this context, however, a contribution of *BGLU30* to GLS degradation is less likely because an increase of *BGLU30* transcripts was observed only after 12 h or longer of dark treatment (Fujiki et al., 2001; Lee J. et al., 2007). In addition to the transcriptomic changes of other *BGLUs*, post-translational regulation of myrosinase activities, including those of TGG1 and TGG2 (Brandt et al., 2018), should be considered in order to gain a better understanding of the dynamic control of GLS contents over a 24-h period.

ADDITIONAL ROLES OF CLASSICAL TGGs BEYOND THE “MUSTARD OIL BOMB”

Several new findings on members of the QE-type myrosinases (TGG1-TGG6) have indicated their broader physiological importance in various situations, beyond the classical intercellular “mustard oil bomb” system deployed against predators. As described above, TGG1 and TGG2 might be involved in guard cell ITC production (Zhao et al., 2008; Islam et al., 2009) and the diurnal control of GLS levels in the absence of tissue disruption (Brandt et al., 2018). In this section we discuss two recent studies that have reported the detailed analysis of root-specific TGGs and the function of TGG6, which had hitherto been considered a pseudogene (Fu et al., 2016a,b).

Currently, QE-type myrosinases are classified into two subclasses, namely, Myr I and Myr II (Wang et al., 2009a). Members of subclass Myr I are found in all GLS-containing plants and are typically deposited in myrosin cells, thereby establishing the compartmentalization required for the intercellular GLS-myrosinase system (Rask et al., 2000). Thus, Myr I class

myrosinases are considered to be critical for biochemical defense against herbivores. Myr II members differ from those in subclass Myr I with respect to several features, including sequence divergence and gene structure (Wang et al., 2009a; Nong et al., 2010; Fu et al., 2016a). Functional analysis of TGG4 and TGG5, the first examples of the Myr II subfamily to be examined, indicated their root-specific roles differ from those of leaf-localized Myr I members. A comparison of the enzymatic properties of TGG4 and TGG5 with those of TGG1 using recombinant proteins expressed in *Pichia pastoris* revealed that TGG4 and TGG5 have higher stability than TGG1 under adverse conditions, such as high temperature, low pH, and excess NaCl (Andersson et al., 2009). In a more recent study, Fu et al. investigated the tissue localization, regulation of root growth, and possible contribution to auxin biosynthesis of TGG4 and TGG5 (Fu et al., 2016b). Analyses of GUS reporter gene expression and myrosinase activities of the single and double KO mutants tended to indicate that TGG5 is more predominant in roots. The defective root elongation under flooded conditions and expression patterns of the auxin-responsive *DR5:GUS* reporter system in these mutants indicated that TGG4 and TGG5 may contribute to auxin biosynthesis at the root tip by hydrolyzing indole GLSs to form indole-3-acetonitrile, a direct precursor of indole-3-acetic acid, even though their enzymatic activities against indole GLSs remain to be confirmed. Given that the aforementioned experiments were performed under non-invasive conditions, it is conceivable that GLS breakdown by the Myr II myrosinases may be less dependent on tissue damage. Hence, not only EE-type but also classical myrosinases could regulate subcellular GLS–myrosinase systems in addition to the so-called “mustard oil bomb.” Our understanding in this regard will be ameliorated by single-cell-level analysis of the specialized compartments in different cell types, as reviewed by Chen et al. (2019) in this issue.

A further surprising finding is that *TGG6*, which had previously been reported to be a pseudogene but is specifically expressed in pollen (Wang et al., 2009b), is still functional in a number of *A. thaliana* accessions (Fu et al., 2016a). The authors identified 10 functional alleles of *TGG6* from 29 accessions and the recombinant *TGG6* derived from Tsu-1 showed a clear myrosinase activity against sinigrin. The predominant expression pattern of functional *TGG6* alleles in pollen was relatively similar to that of the non-functional *TGG6* in Col-0. Given that an ortholog of *TGG6* is predicted to be functional in *Arabidopsis lyrata*, an outcrossing relative of *A. thaliana* (Kusaba et al., 2001; Sherman-Broyles et al., 2007; Tang et al., 2007), it is suggested that its ancestral role was the defense of pollen against herbivores. However, subsequent evolutionary acquisition of a self-fertilization system rendered it no longer critical in *A. thaliana*, thereby resulting in a loss of function in most accessions. A hypothesis proposed based on the findings on *TGG6* is that *BGLUs* with low expression levels in Col-0, the most studied accession of *A. thaliana*, could still be functional in other accessions. As GLS compositions have become significantly differentiated during evolution even within the same species (Kliebenstein et al., 2001; Edgar et al., 2015; Kliebenstein and Cacho, 2016; Barco and Clay, 2019), it is possible that a *BGLU*

plays a critical role in a few accessions but is not important in others. Accession- and species-wide analysis of the same *BGLU* orthologs may help us to gain a better understanding of the specific functions of these enzymes *in planta* and how they have acquired these specific roles during the course of evolution.

CURRENT UNDERSTANDING OF THE OTHER BRASSICALES-SPECIFIC BGLU

Other than *BGLUs* described above, current insights on the Brassicales-specific *BGLUs* are highly limited. To get an overview of *BGLU18*–*BGLU33*, we performed a public data analysis using ePlant¹ (Waese et al., 2017) (Figure 2). In many cases, only a few publications were extracted when each *BGLU* was used as a query (Figure 2A). Our lack of knowledge on several *BGLUs* is probably due to their almost undetectable expression. Using the Plant eFP viewer, we can see that signal levels of these *BGLUs* in the Affymetrix ATH1 microarray are very low in almost all tissues, except for *BGLU18* and *PEN2* (Figure 2B). Although the low signal does not mean a low abundance of the actual transcript, it hinders performance of many biological analyses. Instead, some *BGLUs* exhibit very specific expression patterns in particular tissue(s), e.g., *BGLU19* in mature seeds (Figure 2B and Supplementary Table S1). Moreover, expression levels of each *BGLU* under broad abiotic stresses are summarized (Figure 2C and Supplementary Tables S2, S3). In addition to the known information such as up-regulation of *BGLU18* by drought or wounding, we can expect drastic changes in the expression of uncharacterized *BGLUs* in response to specific stresses, such as the highly increased expression of *BGLU24* in roots as a result of osmotic and salt stress. It should also be noted that according to ATTED-II² (Obayashi et al., 2018), several *BGLUs* show high co-expression scores with specifier proteins: *BGLU19* with *NSP2*, *BGLU30* with *NSP5*, and *PYK10* with *NSP1/NSP3/NSP4* (crosshybridized to the same probe). Broad end products might be generated even in the atypical myrosinase-mediated GLS breakdown. Hence, numerous public data previously collected could help us hypothesize a specific relevance of these *BGLU(s)* in particular developmental stages or abiotic stress responses.

CONCLUDING REMARKS AND FUTURE PERSPECTIVE

In the past decade, subsequent to the identification of *PEN2* in 2009, only *PYK10* has been reported as a further member of the EE-type myrosinases. As reviewed here, however, this does not exclude the possible contribution of Brassicales-specific *BGLUs* to GLS breakdown under specific conditions. In particular, discovery of EE-type myrosinases catalyzing the hydrolysis of aliphatic GLSs in addition to indole GLSs would generate heightened interest amongst researchers in their

¹<https://bar.utoronto.ca/eplant/>

²<http://atted.jp>

physiological importance and catalytic mechanisms, compared with the classical myrosinases. In addition, recent studies have demonstrated that classical myrosinases such as TGGs can also participate in non-tissue-disruptive GLS breakdown beyond the well-known “mustard oil bomb” system at the tissue level. Furthermore, we should pay attention to the accession-wide functional differentiation of the same ortholog with a single species, as highlighted in the case of TGG6. For example, substrate specificity may be dependent on the GLS composition of an accession. It may also be possible that a BGLU has evolved to regulate specialized signals initiated by GLS species present in only a few accessions. Addressing the broad distribution and different myrosinase activities of BGLUs in Brassicaceae and closely related families will enable us to elucidate how the multi-functionality of GLSs is controlled *in planta*, and how the GLS–myrosinase systems have diversified during evolution.

In vitro enzymatic assays using recombinant proteins would be helpful in determining the physiological functions of these enzymes *in planta*. In this regard, *Pichia pastoris* and tobacco BY-2 cells seem to be the preferable organisms to express the *A. thaliana* myrosinases with enzymatic functions, according to previous studies (Andersson et al., 2009; Bednarek et al., 2009; Fu et al., 2016a; Nakano et al., 2017). However, we should bear in mind the fact that the myrosinase assay using sinigrin, an easily available substrate, may not identify (and perhaps has not identified) the actual enzymatic potential of candidate BGLUs of interest. As the classical myrosinases tend to hydrolyze a diverse range of GLS structures (Zhou et al., 2012), most studies have examined the “myrosinase activity” using only sinigrin, a GLS with a simple allyl chain. However, the EE-type myrosinases may have a restricted substrate selectivity and require optimal conditions to work under particular conditions, as emphasized by the findings for PEN2 and PYK10, which preferentially hydrolyze indole GLSs within a narrow optimal pH range (Bednarek et al., 2009; Nakano et al., 2017). It is also notable that sinigrin is detected only in certain *A. thaliana* accessions other than Col-0 (Kliebenstein et al., 2001). Since the side chain structures of GLSs can substantially alter the physicochemical properties of the corresponding degradation products such as ITCs, it would be preferable to examine the activity of myrosinases against a broader range of GLS species to establish the

physiological importance of the BGLUs of interest. Recent advances in the methods for extraction of intact GLSs from plant materials and quantitative analysis of GLS contents may contribute to promoting this approach (Bianco et al., 2017; Doheny-Adams et al., 2017).

In addition to the abiotic stresses discussed herein, there are a few physiological conditions that are potentially related to GLS metabolism and catabolism (Table 2). For example, pre-treatment with phenethyl ITC confers heat stress tolerance on *A. thaliana* seedlings, probably by up-regulating a suite of heat-shock proteins (Hara et al., 2012; Kissen et al., 2016). In addition, the mechanisms underlying the degradation of total GLS amount independent of TGG1 and TGG2 during early developmental stages remain to be clarified (Barth and Jander, 2006). Given that up-regulation of particular BGLUs has yet to be observed under these conditions, we should consider the post-translational regulation of myrosinase activities with regard to myrosinase-associated proteins or small molecule elicitors such as ascorbate (Wittstock et al., 2016a; Bhat and Vyas, 2019; Chen et al., 2019). Under non-disruptive conditions, the physiological functions of myrosinases, including TGGs, are probably controlled more strictly and dynamically than expected till now.

AUTHOR CONTRIBUTIONS

RS and MH prepared the manuscript. RS prepared the figures and tables. MH finalized the manuscript for submission.

FUNDING

This work was supported by the JSPS KAKENHI Grant-in-Aid for Early-Career Scientists (No. 18K14348 to RS) and the RIKEN Special Postdoctoral Researcher Program (to RS).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.01008/full#supplementary-material>

REFERENCES

- Aarabi, F., Kusajima, M., Tohge, T., Konishi, T., Gigolashvili, T., Takamune, M., et al. (2016). Sulfur-deficiency-induced repressor proteins optimize glucosinolate biosynthesis in plants. *Sci. Adv.* 2:e1601087. doi: 10.1126/sciadv.1601087
- Agerbirk, N., Matthes, A., Erthmann, P. O., Ugolini, L., Cinti, S., Lazaridi, E., et al. (2018). Glucosinolate turnover in Brassicales species to an oxazolidin-2-one, formed via the 2-thione and without formation of thioamide. *Phytochemistry* 153, 79–93. doi: 10.1016/j.phytochem.2018.05.006
- Agerbirk, N., and Olsen, C. E. (2012). Glucosinolate structures in evolution. *Phytochemistry* 77, 16–45. doi: 10.1016/j.phytochem.2012.02.005
- Ahn, Y. O., Shimizu, B., Sakata, K., Gantulga, D., Zhou, C., Bevan, D. R., et al. (2010). Scopolin-hydrolyzing beta-glucosidases in roots of *Arabidopsis*. *Plant Cell Physiol.* 51, 132–143. doi: 10.1093/pcp/pcp174
- Andersson, D., Chakrabarty, R., Bejai, S., Zhang, J., Rask, L., and Meijer, J. (2009). Myrosinases from root and leaves of *Arabidopsis thaliana* have different catalytic properties. *Phytochemistry* 70, 1345–1354. doi: 10.1016/j.phytochem.2009.07.036
- Andersson, M. X., Nilsson, A. K., Johansson, O. N., Boztas, G., Adolfsson, L. E., Pinosa, F., et al. (2015). Involvement of the electrophilic isothiocyanate sulforaphane in *Arabidopsis* local defense responses. *Plant Physiol.* 167, 251–261. doi: 10.1104/pp.114.251892
- Banerjee, S., and Paruthy, S. B. (2017). “Therapeutic paradigm underscoring glucosinolate sulforaphane in chemo- and radiosensitization of cancer: preclinical and clinical perspective,” in *Glucosinolates. Reference Series in Phytochemistry*, eds J. M. Mérillon, and K. Ramawat, (Cham: Springer), 339–379. doi: 10.1007/978-3-319-25462-3_19

- Barco, B., and Clay, N. K. (2019). Evolution of glucosinolate diversity via whole-genome duplications, gene rearrangements, and substrate promiscuity. *Annu. Rev. Plant Biol.* 70, 585–604. doi: 10.1146/annurev-arplant-050718-100152
- Barth, C., and Jander, G. (2006). Arabidopsis myrosinases TGG1 and TGG2 have redundant function in glucosinolate breakdown and insect defense. *Plant J.* 46, 549–562. doi: 10.1111/j.1365-313x.2006.02716.x
- Bednarek, P. (2012a). Chemical warfare or modulators of defence responses - the function of secondary metabolites in plant immunity. *Curr. Opin. Plant Biol.* 15, 407–414. doi: 10.1016/j.pbi.2012.03.002
- Bednarek, P. (2012b). Sulfur-containing secondary metabolites from *Arabidopsis thaliana* and other Brassicaceae with function in plant immunity. *ChemBiochem* 13, 1846–1859. doi: 10.1002/cbic.201200086
- Bednarek, P., Pišlewska-Bednarek, M., Svatoš, A., Schneider, B., Doubšký, J., Mansurova, M., et al. (2009). A glucosinolate metabolism pathway in living plant cells mediates broad-spectrum antifungal defense. *Science* 323, 101–106. doi: 10.1126/science.1163732
- Bednarek, P., Pišlewska-Bednarek, M., Ver Loren van Themaat, E., Maddula, R. K., Svatoš, A., and Schulze-Lefert, P. (2011). Conservation and clade-specific diversification of pathogen-inducible tryptophan and indole glucosinolate metabolism in *Arabidopsis thaliana* relatives. *New Phytol.* 192, 713–726. doi: 10.1111/j.1469-8137.2011.03824.x
- Bhat, R., and Vyas, D. (2019). Myrosinase: insights on structural, catalytic, regulatory, and environmental interactions. *Crit. Rev. Biotechnol.* 39, 508–523. doi: 10.1080/07388551.2019.1576024
- Bianco, G., Pascale, R., Lelario, F., Bufo, S. A., and Cataldi, T. R. I. (2017). “Investigation of glucosinolates by mass spectrometry,” in *Glucosinolates. Reference Series in Phytochemistry*, eds J. M. Mérillon, and K. Ramawat, (Cham: Springer), 431–461. doi: 10.1007/978-3-319-25462-3_12
- Brandt, S., Fachinger, S., Tohge, T., Fernie, A. R., Braun, H. P., and Hildebrandt, T. M. (2018). Extended darkness induces internal turnover of glucosinolates in *Arabidopsis thaliana* leaves. *PLoS One* 13:e0202153. doi: 10.1371/journal.pone.0202153
- Brown, P. D., Tokuhisa, J. G., Reichelt, M., and Gershenzon, J. (2003). Variation of glucosinolate accumulation among different organs and developmental stages of *Arabidopsis thaliana*. *Phytochemistry* 62, 471–481. doi: 10.1016/s0031-9422(02)00549-6
- Burmeister, W. P., Cottaz, S., Driguez, H., Iori, R., Palmieri, S., and Henrissat, B. (1997). The crystal structures of Sinapis alba myrosinase and a covalent glycosyl-enzyme intermediate provide insights into the substrate recognition and active-site machinery of an S-glycosidase. *Structure* 5, 663–675.
- Burrow, M. (2016). “Complex environments interact with plant development to shape glucosinolate profiles,” in *Advances in Botanical Research*, Vol. 80, ed. S. Kopriva, (Cambridge, MA: Academic Press), 15–30. doi: 10.1016/bs.abr.2016.06.001
- Burrow, M., and Halkier, B. A. (2017). How does a plant orchestrate defense in time and space? Using glucosinolates in Arabidopsis as case study. *Curr. Opin. Plant Biol.* 38, 142–147. doi: 10.1016/j.pbi.2017.04.009
- Burrow, M., Halkier, B. A., and Kliebenstein, D. J. (2010). Regulatory networks of glucosinolates shape *Arabidopsis thaliana* fitness. *Curr. Opin. Plant Biol.* 13, 347–352. doi: 10.1016/j.pbi.2010.02.002
- Buwalda, F., De Kok, L. J., Stulen, I., and Kuiper, P. J. C. (1988). Cysteine, γ -glutamylcysteine and glutathione contents of spinach leaves as affected by darkness and application of excess sulfur. *Physiol. Plant.* 74, 663–668. doi: 10.1111/j.1399-3054.1988.tb02033.x
- Cao, Y.-Y., Yang, J.-F., Liu, T.-Y., Su, Z.-F., Zhu, F.-Y., Chen, M.-X., et al. (2017). A phylogenetically informed comparison of GH1 hydrolases between Arabidopsis and rice response to stressors. *Front. Plant Sci.* 8:350. doi: 10.3389/fpls.2017.00350
- Castro, A., Stulen, I., and De Kok, L. J. (2003). “Nitrogen and sulfur requirement of Brassica oleracea L. cultivars,” in *Sulfur Transport and Assimilation in Plants*, eds J. C. Davidian, D. Grill, L. J. De Kok, M. J. Stulen, E. S. Hawkesford, et al. (Leiden: Backhuys Publishers), 181–183.
- Chen, S., Chhajed, S., Misra, B. B., and Tello, N. (2019). Chemodiversity of the glucosinolate-myrosinase system at the single cell-type resolution. *Front. Plant Sci.* 10:618. doi: 10.3389/fpls.2019.00618
- Clay, N. K., Adio, A. M., Denoux, C., Jander, G., and Ausubel, F. M. (2009). Glucosinolate metabolites required for an Arabidopsis innate immune response. *Science* 323, 95–101. doi: 10.1126/science.1164627
- Daszkowska-Golec, A. (2016). “The role of abscisic acid in drought stress: how ABA helps plants to cope with drought stress,” in *Drought Stress Tolerance in Plants*, Vol. 2, eds M. Hossain, S. Wani, S. Bhattacharjee, D. Burritt, and L. S. Tran, (Cham: Springer), 123–151. doi: 10.1007/978-3-319-32423-4_5
- Doheny-Adams, T., Redeker, K., Kittipol, V., Bancroft, I., and Hartley, S. E. (2017). Development of an efficient glucosinolate extraction method. *Plant Methods* 13:17. doi: 10.1186/s13007-017-0164-168
- Edger, P. P., Heideel-Fischer, H. M., Bekaert, M., Rota, J., Glöckner, G., Platts, A. E., et al. (2015). The butterfly plant arms-race escalated by gene and genome duplications. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8362–8366. doi: 10.1073/pnas.1503926112
- Eisenschmidt-Bönn, D., Schneegans, N., Backenköhler, A., Wittstock, U., and Brandt, W. (2019). Structural diversification during glucosinolate breakdown: mechanisms of thiocyanate, epithionitrile and simple nitrile formation. *Plant J.* 99, 329–343. doi: 10.1111/tpj.14327
- Fahey, J. W., Zalcmann, A. T., and Talalay, P. (2001). The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. *Phytochemistry* 56, 5–51. doi: 10.1016/s0031-9422(00)00316-2
- Falk, K. L., Tokuhisa, J. G., and Gershenzon, J. (2007). The effect of sulfur nutrition on plant glucosinolate content: physiology and molecular mechanisms. *Plant Biol.* 9, 573–581. doi: 10.1055/s-2007-965431
- Francisco, M., Joseph, B., Caligagan, H., Li, B., Corwin, J. A., Lin, C., et al. (2016a). Genome wide association mapping in *Arabidopsis thaliana* identifies novel genes involved in linking allyl glucosinolate to altered biomass and defense. *Front. Plant Sci.* 7:1010. doi: 10.3389/fpls.2016.01010
- Francisco, M., Joseph, B., Caligagan, H., Li, B., Corwin, J. A., Lin, C., et al. (2016b). The defense metabolite, allyl glucosinolate, modulates *Arabidopsis thaliana* biomass dependent upon the endogenous glucosinolate pathway. *Front. Plant Sci.* 7:774. doi: 10.3389/fpls.2016.00774
- Frerigmann, H. (2016). “Glucosinolate regulation in a complex relationship – MYC and MYB – No one can act without each other,” in *Advances in Botanical Research*, Vol. 80, ed. S. Kopriva, (Cambridge, MA: Academic Press), 57–97. doi: 10.1016/bs.abr.2016.06.005
- Frerigmann, H., Pišlewska-Bednarek, M., Sánchez-Vallet, A., Molina, A., Glawischnig, E., Gigolashvili, T., et al. (2016). Regulation of pathogen-triggered tryptophan metabolism in *Arabidopsis thaliana* by MYB transcription factors and indole glucosinolate conversion products. *Mol. Plant* 9, 682–695. doi: 10.1016/j.molp.2016.01.006
- Fu, L., Han, B., Tan, D., Wang, M., Ding, M., and Zhang, J. (2016a). Identification and evolution of functional alleles of the previously described pollen specific myrosinase pseudogene AtTGG6 in *Arabidopsis thaliana*. *Int. J. Mol. Sci.* 17:262. doi: 10.3390/ijms17020262
- Fu, L., Wang, M., Han, B., Tan, D., Sun, X., and Zhang, J. (2016b). Arabidopsis myrosinase genes AtTGG4 and AtTGG5 are root-tip specific and contribute to auxin biosynthesis and root-growth regulation. *Int. J. Mol. Sci.* 17:892. doi: 10.3390/ijms17060892
- Fujiki, Y., Yoshikawa, Y., Sato, T., Inada, N., Ito, M., Nishida, I., et al. (2001). Dark-inducible genes from *Arabidopsis thaliana* are associated with leaf senescence and repressed by sugars. *Physiol. Plant.* 111, 345–352. doi: 10.1034/j.1399-3054.2001.1110312.x
- Halkier, B. A. (2016). “General introduction to glucosinolates,” in *Advances in Botanical Research*, Vol. 80, ed. S. Kopriva, (Cambridge, MA: Academic Press), 1–14. doi: 10.1016/bs.abr.2016.07.001
- Halkier, B. A., and Gershenzon, J. (2006). Biology and biochemistry of glucosinolates. *Annu. Rev. Plant Biol.* 57, 303–333. doi: 10.1146/annurev-arplant.57.032905.105228
- Hansch, F. S., Evelyn Lamy, E., Schreiner, M., and Rohn, S. (2014). Reactivity and stability of glucosinolates and their breakdown products in foods. *Angew. Chem. Int. Ed.* 53, 11430–11450. doi: 10.1002/anie.201402639
- Hara, M., Harazaki, A., and Tabata, K. (2012). Administration of isothiocyanates enhances heat tolerance in *Arabidopsis thaliana*. *Plant Growth Regul.* 69, 71–77. doi: 10.1007/s10725-012-9748-5
- Haughn, G. W., Davin, L., Giblin, M., and Underhill, E. W. (1991). Biochemical genetics of plant secondary metabolites in *Arabidopsis thaliana*. *Plant Physiol.* 97, 217–226. doi: 10.1104/pp.97.1.217
- Hirai, M. Y., Fujiwara, T., Awazu, M., Kimura, T., Noji, M., and Saito, K. (2003). Global expression profiling of sulfur-starved *Arabidopsis* by DNA microarray reveals the role of O-acetyl-L-serine as a general regulator of gene

- expression in response to sulfur nutrition. *Plant J.* 33, 651–663. doi: 10.1046/j.1365-3113.2003.01658.x
- Hirai, M. Y., Klein, M., Fujikawa, Y., Yano, M., Goodenowe, D. B., Yamazaki, Y., et al. (2005). Elucidation of gene-to-gene and metabolite-to-gene networks in *Arabidopsis* by integration of metabolomics and transcriptomics. *J. Biol. Chem.* 280, 25590–25595. doi: 10.1074/jbc.m502332200
- Hirai, M. Y., and Saito, K. (2004). Post-genomics approaches for the elucidation of plant adaptive mechanisms to sulphur deficiency. *J. Exp. Bot.* 55, 1871–1879. doi: 10.1093/jxb/erh184
- Hirai, M. Y., Yano, M., Goodenowe, D. B., Kanaya, S., Kimura, T., Awazuhara, M., et al. (2004). Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 101, 10205–10210. doi: 10.1073/pnas.0403218101
- Hiruma, K., Onozawa-Komori, M., Takahashi, F., Asakura, M., Bednarek, P., Okuno, T., et al. (2010). Entry mode-dependent function of an indole glucosinolate pathway in *Arabidopsis* for nonhost resistance against anthracnose pathogens. *Plant Cell* 22, 2429–2443. doi: 10.1105/tpc.110.074344
- Hossain, M. S., Ye, W., Hossain, M. A., Okuma, E., Uraji, M., Nakamura, Y., et al. (2013). Glucosinolate degradation products, isothiocyanates, nitriles, and thiocyanates, induce stomatal closure accompanied by peroxidase-mediated reactive oxygen species production in *Arabidopsis thaliana*. *Biosci. Biotechnol. Biochem.* 77, 977–983. doi: 10.1271/bbb.120928
- Huseby, S., Koprivova, A., Lee, B.-R., Saha, S., Mithen, R., Wold, A.-B., et al. (2013). Diurnal and light regulation of sulphur assimilation and glucosinolate biosynthesis in *Arabidopsis*. *J. Exp. Bot.* 64, 1039–1048. doi: 10.1093/jxb/ers378
- Islam, M. M., Tani, C., Watanabe-Sugimoto, M., Uraji, M., Jahan, M. S., Masuda, C., et al. (2009). Myrosinases, TGG1 and TGG2, redundantly function in ABA and MeJA signaling in *Arabidopsis* guard cells. *Plant Cell Physiol.* 50, 1171–1175. doi: 10.1093/pcp/pcp066
- Jackson, T. L., Baker, G. W., Wilks, F. R. Jr., Popov, V. A., Mathur, J., and Benfey, P. N. (2015). Large cellular inclusions accumulate in *Arabidopsis* roots exposed to low-sulfur conditions. *Plant Physiol.* 168, 1573–1589. doi: 10.1104/pp.15.00465
- Johansson, O. N., Fantozzi, E., Fahlberg, P., Nilsson, A. K., Buhot, N., Tör, M., et al. (2014). Role of the penetration-resistance genes PEN1, PEN2 and PEN3 in the hypersensitive response and race-specific resistance in *Arabidopsis thaliana*. *Plant J.* 79, 466–476. doi: 10.1111/tpj.12571
- Jørgensen, M. E., Nour-Eldin, H. H., and Halkier, B. A. (2015). Transport of defense compounds from source to sink: lessons learned from glucosinolates. *Trends Plant Sci.* 20, 508–514. doi: 10.1016/j.tplants.2015.04.006
- Katz, E., and Chamovitz, D. A. (2017). Wounding of *Arabidopsis* leaves induces indole-3-carbinol-dependent autophagy in roots of *Arabidopsis thaliana*. *Plant J.* 91, 779–787. doi: 10.1111/tpj.13610
- Katz, E., Nisani, S., Yadav, B. S., Woldemariam, M. G., Shai, B., Obolski, U., et al. (2015). The glucosinolate breakdown product indole-3-carbinol acts as an auxin antagonist in roots of *Arabidopsis thaliana*. *Plant J.* 82, 547–555. doi: 10.1111/tpj.12824
- Keling, H., and Zhu, Z. (2010). Effects of different concentrations of sodium chloride on plant growth and glucosinolate content and composition in pakchoi. *Afr. J. Biotechnol.* 9, 4428–4433.
- Khokon, A. R., Jahan, S., Rahman, T., Hossain, M. A., Muroyama, D., Minami, I., et al. (2011). Allyl isothiocyanate (AITC) induces stomatal closure in *Arabidopsis*. *Plant Cell Environ.* 34, 1900–1906. doi: 10.1111/j.1365-3040.2011.02385.x
- Kilian, J., Whitehead, D., Horak, J., Wanke, D., Weinl, S., Batistic, O., et al. (2007). The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. *Plant J.* 50, 347–363. doi: 10.1111/j.1365-3113.2007.03052.x
- Kissen, R., Øverby, A., Winge, P., and Bones, A. M. (2016). Allyl-isothiocyanate treatment induces a complex transcriptional reprogramming including heat stress, oxidative stress and plant defence responses in *Arabidopsis thaliana*. *BMC Genomics* 17:740. doi: 10.1186/s12864-016-3039-x
- Klein, A. P., and Sattely, E. S. (2015). Two cytochromes P450 catalyze S-heterocyclizations in cabbage phytoalexin biosynthesis. *Nat. Chem. Biol.* 11, 837–839. doi: 10.1038/nchembio.1914
- Klein, A. P., and Sattely, E. S. (2017). Biosynthesis of cabbage phytoalexins from indole glucosinolate. *Proc. Natl. Acad. Sci. U.S.A.* 114, 1910–1915. doi: 10.1073/pnas.1615625114
- Klein, M., Reichelt, M., Gershenzon, J., and Papenbrock, J. (2006). The three desulfoglucosinolate sulfotransferase proteins in *Arabidopsis* have different substrate specificities and are differentially expressed. *FEBS J.* 273, 122–136. doi: 10.1111/j.1742-4658.2005.05048.x
- Kliebenstein, D. J., and Cacho, N. I. (2016). “Nonlinear selection and a blend of convergent, divergent and parallel evolution shapes natural variation in glucosinolates,” in *Advances in Botanical Research*, Vol. 80, ed. S. Kopriva, (Cambridge, MA: Academic Press), 31–55. doi: 10.1016/bs.abr.2016.06.002
- Kliebenstein, D. J., Kroymann, J., Brown, P., Figuth, A., Pedersen, D., Gershenzon, J., et al. (2001). Genetic control of natural variation in *Arabidopsis* glucosinolate accumulation. *Plant Physiol.* 126, 811–825. doi: 10.1104/pp.126.2.811
- Kocsy, G., Owttrim, G., Brander, K., and Brunold, C. (1997). Effect of chilling on the diurnal rhythm of enzymes involved in protection against oxidative stress in a chilling-tolerant and a chilling-sensitive maize genotype. *Physiol. Plant.* 99, 249–254. doi: 10.1034/j.1399-3054.1997.990206.x
- Kopriva, S. (ed.) (2016). *Advances in Botanical Research*, Vol. 80. Amsterdam: Academic Press.
- Kopriva, S., Muheim, R., Koprivova, A., Trachsel, N., Catalano, C., Suter, M., et al. (1999). Light regulation of assimilatory sulphate reduction in *Arabidopsis thaliana*. *Plant J.* 20, 37–44. doi: 10.1046/j.1365-3113.1999.00573.x
- Koroleva, O. A., and Cramer, R. (2011). Single-cell proteomic analysis of glucosinolate-rich S-cells in *Arabidopsis thaliana*. *Methods* 54, 413–423. doi: 10.1016/j.jymeth.2011.06.005
- Kusaba, M., Dwyer, K., Hendershot, J., Vrebalov, J., Nasrallah, J. B., and Nasrallah, M. E. (2001). Self-incompatibility in the genus *Arabidopsis*: characterization of the S locus in the outcrossing *A. lyrata* and its autogamous relative *A. thaliana*. *Plant Cell* 13, 627–643. doi: 10.1105/tpc.13.3.627
- Lambrix, V., Reichelt, M., Mitchell-Olds, T., Kliebenstein, D. J., and Gershenzon, J. (2001). The *Arabidopsis* epithiospecifier protein promotes the hydrolysis of glucosinolates to nitriles and influences *Trichoplusia ni* herbivory. *Plant Cell* 13, 2793–2807. doi: 10.1105/tpc.13.12.2793
- Lee, B.-R., Koprivova, A., and Kopriva, S. (2011). The key enzyme of sulfate assimilation, adenosine 5'-phosphosulfate reductase, is regulated by HY5 in *Arabidopsis*. *Plant J.* 67, 1042–1054. doi: 10.1111/j.1365-3113.2011.04656.x
- Lee, E.-J., Matsumura, Y., Soga, K., Hoson, T., and Koizumi, N. (2007). Glycosyl hydrolases of cell wall are induced by sugar starvation in *Arabidopsis*. *Plant Cell Physiol.* 48, 405–413. doi: 10.1093/pcp/pcm009
- Lee, J., He, K., Stolc, V., Lee, H., Figueroa, P., Gao, Y., et al. (2007). Analysis of transcription factor HY5 genomic binding sites revealed its hierarchical role in light regulation of development. *Plant Cell* 19, 731–749. doi: 10.1105/tpc.106.047688
- Lee, K. H., Piao, H. L., Kim, H. Y., Choi, S. M., Jiang, F., Hartung, W., et al. (2006). Activation of glucosidase via stress-induced polymerization rapidly increases active pools of abscisic acid. *Cell* 126, 1109–1120. doi: 10.1016/j.cell.2006.07.034
- Lipka, V., Dittgen, J., Bednarek, P., Bhat, R., Wiermer, M., Stein, M., et al. (2005). Pre- and postinvasion defenses both contribute to nonhost resistance in *Arabidopsis*. *Science* 310, 1180–1183. doi: 10.1126/science.1119409
- López-Berenguer, C., Martínez-Ballesta, M. C., García-Viguera, C., and Carvajal, M. (2008). Leaf water balance mediated by aquaporins under salt stress and associated glucosinolate synthesis in broccoli. *Plant Sci.* 174, 321–328. doi: 10.1016/j.plantsci.2007.11.012
- Lu, X., Dittgen, J., Piślewska-Bednarek, M., Molina, A., Schneider, B., Svatos, A., et al. (2015). Mutant allele-specific uncoupling of PENETRATION3 functions reveals engagement of the ATP-binding cassette transporter in distinct tryptophan metabolic pathways. *Plant Physiol.* 168, 814–827. doi: 10.1104/pp.15.00182
- Ludwig-Müller, J., Krishna, P., and Forreiter, C. (2000). A glucosinolate mutant of *Arabidopsis* is thermosensitive and defective in cytosolic Hsp90 expression after heat stress. *Plant Physiol.* 123, 949–958. doi: 10.1104/pp.123.3.949
- Lüthy, B., and Matile, P. (1984). The mustard oil bomb: rectified analysis of the subcellular organisation of the myrosinase system. *Biochem. Physiol. Pflanzen* 179, 5–12. doi: 10.1016/s0015-3796(84)80059-1
- Maeda, K., Houjyou, Y., Komatsu, T., Hori, H., Kodaira, T., and Ishikawa, A. (2009). AGB1 and PMR5 contribute to PEN2-mediated preinvasion resistance

- to *Magnaporthe oryzae* in *Arabidopsis thaliana*. *Mol. Plant Microbe Interact.* 22, 1331–1340. doi: 10.1094/MPMI-22-11-1331
- Malinovsky, F. G., Thomsen, M.-L. F., Nintemann, S. J., Jagd, L. M., Bourguine, B., Burow, M., et al. (2017). An evolutionarily young defense metabolite influences the root growth of plants via the ancient TOR signaling pathway. *eLife* 6:e29353. doi: 10.7554/eLife.29353
- Malka, S. K., and Cheng, Y. (2017). Possible interactions between the biosynthetic pathways of indole glucosinolate and auxin. *Front. Plant Sci.* 8:2131. doi: 10.3389/fpls.2017.02131
- Martinez-Ballesta, M. C., Moreno, D. A., and Carvajal, M. (2013). The physiological importance of glucosinolates on plant response to abiotic stress in Brassica. *Int. J. Mol. Sci.* 14, 11607–11625. doi: 10.3390/ijms140611607
- Martinez-Ballesta, M. C., Moreno-Fernández, D. A., Castejon, D., Ochando, C., Morandini, P. A., and Carvajal, M. (2015). The impact of the absence of aliphatic glucosinolates on water transport under salt stress in *Arabidopsis thaliana*. *Front. Plant Sci.* 6:524. doi: 10.3389/fpls.2015.00524
- Maruyama-Nakashita, A., Inoue, E., Watanabe-Takahashi, A., Yamaya, T., and Takahashi, H. (2003). Transcriptome profiling of sulfur-responsive genes in *Arabidopsis* reveals global effects of sulfur nutrition on multiple metabolic pathways. *Plant Physiol.* 132, 597–605. doi: 10.1104/pp.102.019802
- Maruyama-Nakashita, A., Nakamura, Y., Tohge, T., Saito, K., and Takahashi, H. (2006). *Arabidopsis* SLIM1 is a central transcriptional regulator of plant sulfur response and metabolism. *Plant Cell* 18, 3235–3251. doi: 10.1105/tpc.106.046458
- Matsushima, R., Hayashi, Y., Kondo, M., Shimada, T., Nishimura, M., and Hara-Nishimura, I. (2002). An endoplasmic reticulum-derived structure that is induced under stress conditions in *Arabidopsis*. *Plant Physiol.* 130, 1807–1814. doi: 10.1104/pp.009464
- Matsushima, R., Kondo, M., Nishimura, M., and Hara-Nishimura, I. (2003). A novel ER-derived compartment, the ER body, selectively accumulates a beta-glucosidase with an ER-retention signal in *Arabidopsis*. *Plant J.* 33, 493–502. doi: 10.1046/j.1365-3113x.2003.01636.x
- Mérillon, J. M., and Ramawat, K. (eds) (2017). *Reference Series in Phytochemistry*. Cham: Springer.
- Merilo, E., Jalakas, P., Kollist, H., and Brosché, M. (2015). The role of ABA recycling and transporter proteins in rapid stomatal responses to reduced air humidity, elevated CO₂, and exogenous ABA. *Mol. Plant* 8, 657–659. doi: 10.1016/j.molp.2015.01.014
- Murata, Y., Mori, I. C., and Munemasa, S. (2015). Diverse stomatal signaling and the signal integration mechanism. *Annu. Rev. Plant Biol.* 66, 369–392. doi: 10.1146/annurev-arplant-043014-114707
- Nagano, A. J., Fukao, Y., Fujiwara, M., Nishimura, M., and Hara-Nishimura, I. (2008). Antagonistic jacalin-related lectins regulate the size of ER body-type beta-glucosidase complexes in *Arabidopsis thaliana*. *Plant Cell Physiol.* 49, 969–980. doi: 10.1093/pcp/pcn075
- Nakabayashi, K., Okamoto, M., Koshihara, T., Kamiya, Y., and Nambara, E. (2005). Genome-wide profiling of stored mRNA in *Arabidopsis thaliana* seed germination: epigenetic and genetic regulation of transcription in seed. *Plant J.* 41, 697–709. doi: 10.1111/j.1365-3113x.2005.02337.x
- Nakano, R. T., Piślewska-Bednarek, M., Yamada, K., Edger, P. P., Miyahara, M., Kondo, M., et al. (2017). PYK10 myrosinase reveals a functional coordination between endoplasmic reticulum bodies and glucosinolates in *Arabidopsis thaliana*. *Plant J.* 89, 204–220. doi: 10.1111/tpj.13377
- Nakano, R. T., Yamada, K., Bednarek, P., Nishimura, M., and Hara-Nishimura, I. (2014). ER bodies in plants of the Brassicales order: biogenesis and association with innate immunity. *Front. Plant Sci.* 5:73. doi: 10.3389/fpls.2014.00073
- Nakazaki, A., Yamada, K., Kunieda, T., Sugiyama, R., Hirai, M. Y., Tamura, K., et al. (2019). Leaf endoplasmic reticulum bodies identified in *Arabidopsis* rosette leaves are involved in defense against herbivory. *Plant Physiol.* 179, 1515–1524. doi: 10.1104/pp.18.00984
- Nambara, E., and Marion-Poll, A. (2005). Abscisic acid biosynthesis and catabolism. *Annu. Rev. Plant Biol.* 56, 165–185. doi: 10.1146/annurev.arplant.56.032604.144046
- Nikiforova, V., Freitag, J., Kempa, S., Adamik, M., Hesse, H., and Hoefgen, R. (2003). Transcriptome analysis of sulfur depletion in *Arabidopsis thaliana*: interlacing of biosynthetic pathways provides response specificity. *Plant J.* 33, 633–650. doi: 10.1046/j.1365-3113x.2003.01657.x
- Nintemann, S. J., Hunziker, P., Andersen, T. G., Schulz, A., Burow, M., and Halkier, B. A. (2018). Localization of the glucosinolate biosynthetic enzymes reveals distinct spatial patterns for the biosynthesis of indole and aliphatic glucosinolates. *Physiol. Plant.* 163, 138–154. doi: 10.1111/ppl.12672
- Nong, H., Zhang, J.-M., Li, D., Wang, M., Sun, X., Zhu, Y. J., et al. (2010). Characterization of a novel beta-thioglucosidase CpTGG1 in *Carica papaya* and its substrate-dependent and ascorbic acid-independent O-β-glucosidase activity. *J. Integr. Plant Biol.* 52, 879–890. doi: 10.1111/j.1744-7909.2010.00988.x
- Nour-Eldin, H. H., Andersen, T. G., Burow, M., Madsen, S. R., Jorgensen, M. E., Olsen, C. E., et al. (2012). NRT/PTR transporters are essential for translocation of glucosinolate defence compounds to seeds. *Nature* 488, 531–534. doi: 10.1038/nature11285
- Obayashi, T., Aoki, Y., Tadaka, S., Kagaya, Y., and Kinoshita, K. (2018). ATTED-II in 2018: a plant coexpression database based on investigation of the statistical property of the mutual rank index. *Plant Cell Physiol.* 59:e3. doi: 10.1093/pcp/pcx191
- Ogasawara, K., Yamada, K., Christeller, J. T., Kondo, M., Hatsugai, N., Hara-Nishimura, I., et al. (2009). Constitutive and inducible ER bodies of *Arabidopsis thaliana* accumulate distinct beta-glucosidases. *Plant Cell Physiol.* 50, 480–488. doi: 10.1093/pcp/pcp007
- Ondzighi-Assoume, C. A., Chakraborty, S., and Harris, J. M. (2016). Environmental nitrate stimulates abscisic acid accumulation in *Arabidopsis* root tips by releasing it from inactive stores. *Plant Cell* 28, 729–745. doi: 10.1105/tpc.15.00946
- Passera, C., Ghisi, R., and Ferretti, M. (1989). Light-activation of ATPsulfurylase in leaves and chloroplasts of *Zea mays*. *Photosynthetica* 23, 166–172.
- Pastorczyk, M., and Bednarek, P. (2016). “The function of glucosinolates and related metabolites in plant innate immunity,” in *Advances in Botanical Research*, Vol. 80, ed. S. Kopriva, (Cambridge, MA: Academic Press), 171–198. doi: 10.1016/bs.abr.2016.06.007
- Pedras, M. S. C., Yaya, E. E., and Glawischnig, E. (2011). The phytoalexins from cultivated and wild crucifers: chemistry and biology. *Nat. Prod. Rep.* 28, 1381–1405. doi: 10.1039/c1np00020a
- Piślewska-Bednarek, M., Nakano, R. T., Hiruma, K., Pastorczyk, M., Sánchez-Vallet, A., Singkaravanit-Ogawa, S., et al. (2018). Glutathione transferase U13 functions in pathogen-triggered glucosinolate metabolism. *Plant Physiol.* 176, 538–551. doi: 10.1104/pp.17.01455
- Possenti, M., Baima, S., Raffo, A., Durazzo, A., Giusti, A. M., and Natella, F. (2017). “Glucosinolates in food,” in *Glucosinolates. Reference Series in Phytochemistry*, eds J. M. Mérillon, and K. Ramawat, (Cham: Springer), 87–132.
- Pruneda-Paz, J. L., and Kay, S. A. (2010). An expanding universe of circadian networks in higher plants. *Trends Plant Sci.* 15, 259–265. doi: 10.1016/j.tplants.2010.03.003
- Rask, L., Andréasson, E., Ekblom, B., Eriksson, S., Pontoppidan, B., and Meijer, J. (2000). Myrosinase: gene family evolution and herbivore defense in Brassicaceae. *Plant Mol. Biol.* 42, 93–113. doi: 10.1007/978-94-011-4221-2_5
- Ren, H., Zhong, H. X., Dai, S. J., Chen, S. X., and Yan, X. F. (2009). Water stress on glucosinolate contents in *Arabidopsis* rosette leaves. *Acta Ecol. Sin.* 8, 4372–4379.
- Rosa, E. A. S. (1997). Daily variation in glucosinolate concentrations in the leaves and roots of cabbage seedlings in two constant temperature regimes. *J. Sci. Food Agric.* 73, 364–368. doi: 10.1002/(sici)1097-0010(199703)73:3<364::aid-jfsa742>3.0.co;2-o
- Sakata, Y., Komatsu, K., and Takezawa, D. (2014). “ABA as a Universal Plant Hormone,” in *Progress in Botany*, Vol. 75, eds U. Lüttge, W. Beyschlag, and J. Cushman, (Berlin: Springer), 57–96. doi: 10.1007/978-3-642-38797-5_2
- Schmid, M., Davison, T. S., Henz, S. R., Pape, U. J., Demar, M., Vingron, M., et al. (2005). A gene expression map of *Arabidopsis thaliana* development. *Nat. Genet.* 37, 501–506.
- Schnug, E., and Haneklaus, S. (2016). “Glucosinolates—the agricultural story,” in *Advances in Botanical Research*, Vol. 80, ed. S. Kopriva, (Cambridge, MA: Academic Press), 281–302. doi: 10.1016/bs.abr.2016.07.003
- Schuster, J., Knill, T., Reichelt, M., Gershenzon, J., and Binder, S. (2006). Branched-chain aminotransferase4 is part of the chain elongation pathway in the biosynthesis of methionine-derived glucosinolates in *Arabidopsis*. *Plant Cell* 18, 2664–2679. doi: 10.1105/tpc.105.039339

- Sherman-Broyles, S., Boggs, N., Farkas, A., Liu, P., Vrebalov, J., Nasrallah, M. E., et al. (2007). S locus genes and the evolution of self-fertility in *Arabidopsis thaliana*. *Plant Cell* 19, 94–106. doi: 10.1105/tpc.106.048199
- Shirakawa, M., and Hara-Nishimura, I. (2018). Specialized vacuoles of myrosin cells: chemical defense strategy in Brassicales plants. *Plant Cell Physiol.* 59, 1309–1316. doi: 10.1093/pcp/pcy082
- Tang, C., Toomajian, C., Sherman-Broyles, S., Plagnol, V., Guo, Y.-L., Hu, T. T., et al. (2007). The evolution of selfing in *Arabidopsis thaliana*. *Science* 317, 1070–1072.
- Traka, M. H. (2016). “Health benefits of glucosinolates,” in *Advances in Botanical Research*, Vol. 80, ed. S. Kopriva, (Cambridge, MA: Academic Press), 247–279. doi: 10.1016/bs.abr.2016.06.004
- Urbancsok, J., Bones, A. M., and Kissen, R. (2017). Glucosinolate-derived isothiocyanates inhibit *Arabidopsis* growth and the potency depends on their side chain structure. *Int. J. Mol. Sci.* 18:2372. doi: 10.3390/ijms18112372
- Urbancsok, J., Bones, A. M., and Kissen, R. (2018a). *Arabidopsis* mutants impaired in glutathione biosynthesis exhibit higher sensitivity towards the glucosinolate hydrolysis product allyl-isothiocyanate. *Sci. Rep.* 8:8909. doi: 10.1038/s41598-018-28099-1
- Urbancsok, J., Bones, A. M., and Kissen, R. (2018b). Benzyl cyanide leads to auxin-like effects through the action of nitrilases in *Arabidopsis thaliana*. *Front. Plant Sci.* 9:1240. doi: 10.3389/fpls.2018.01240
- Variyar, P. S. A., Banerjee, A., Akkarakaran, J. J., and Suprasanna, P. (2014). “Role of glucosinolates in plant stress tolerance,” in *Emerging Technologies and Management of Crop Stress Tolerance*, Vol. 1, eds P. Ahmad, and S. Rasool, (Cambridge, MA: Academic Press), 271–291. doi: 10.1016/b978-0-12-800876-8.00012-6
- Vik, D., Mitarai, N., Wulff, N., Halkier, B. A., and Burow, M. (2018). Dynamic modeling of indole glucosinolate hydrolysis and its impact on auxin signaling. *Front. Plant Sci.* 9:550. doi: 10.3389/fpls.2018.00550
- Waese, J., Fan, J., Pasha, A., Yu, H., Fucile, G., Shi, R., et al. (2017). ePlant: visualizing and exploring multiple levels of data for hypothesis generation in plant biology. *Plant Cell* 29, 1806–1821. doi: 10.1105/tpc.17.00073
- Walker, K. C., and Booth, E. J. (2003). “Sulphur nutrition and oilseed quality,” in *Sulphur in Plants*, eds Y. P. Abrol, and A. Ahmad, (Dordrecht: Kluwer Academic Publishers), 323–339. doi: 10.1007/978-94-017-0289-8_18
- Wang, M., Li, D., Sun, X., Zhu, Y. J., Nong, H., and Zhang, J. (2009a). Characterization of a root-specific beta-thioglucoside glucosylhydrolase gene in *Carica papaya* and its recombinant protein expressed in *Pichia pastoris*. *Plant Sci.* 177, 716–723. doi: 10.1016/j.plantsci.2009.09.013
- Wang, M., Sun, X., Tan, D., Gong, S., Meijer, J., and Zhang, J. (2009b). The two non-functional myrosinase genes TGG3 and TGG6 in *Arabidopsis* are expressed predominantly in pollen. *Plant Sci.* 177, 371–375. doi: 10.1016/j.plantsci.2009.06.003
- Wittstock, U., Kurzbach, E., Herfurth, A.-M., and Stauber, E. J. (2016a). “Glucosinolate breakdown,” in *Advances in Botanical Research*, Vol. 80, ed. S. Kopriva, (Cambridge, MA: Academic Press), 125–169. doi: 10.1016/bs.abr.2016.06.006
- Wittstock, U., Meier, K., Dörr, F., and Ravindran, B. M. (2016b). NSP-dependent simple nitrile formation dominates upon breakdown of major aliphatic glucosinolates in roots, seeds, and seedlings of *Arabidopsis thaliana* Columbia-0. *Front. Plant Sci.* 7:1821. doi: 10.3389/fpls.2016.01821
- Xu, J., Meng, J., Meng, X., Zhao, Y., Liu, J., Sun, T., et al. (2016). Pathogen-responsive MPK3 and MPK6 reprogram the biosynthesis of indole glucosinolates and their derivatives in *Arabidopsis* immunity. *Plant Cell* 28, 1144–1162. doi: 10.1105/tpc.15.00871
- Xu, Z., Escamilla-Treviño, L. L., Zeng, L., Lalgondar, M., Bevan, D. R., Winkel, B. S. J., et al. (2004). Functional genomic analysis of *Arabidopsis thaliana* glycoside hydrolase family 1. *Plant Mol. Biol.* 55, 343–367. doi: 10.1007/s11103-004-0790-1
- Xu, Z.-J., Nakajima, M., Suzuki, Y., and Yamaguchi, I. (2002). Cloning and characterization of the abscisic acid-specific glucosyltransferase gene from adzuki bean seedlings. *Plant Physiol.* 129, 1285–1295. doi: 10.1104/pp.001784
- Xu, Z.-Y., Lee, K. H., Dong, T., Jeong, J. C., Jin, J. B., Kanno, Y., et al. (2012). A vacuolar beta-glucosidase homolog that possesses glucose-conjugated abscisic acid hydrolyzing activity plays an important role in osmotic stress responses in *Arabidopsis*. *Plant Cell* 24, 2184–2199. doi: 10.1105/tpc.112.095935
- Yamada, K., Hara-Nishimura, I., and Nishimura, M. (2011). Unique defense strategy by the endoplasmic reticulum body in plants. *Plant Cell Physiol.* 52, 2039–2049. doi: 10.1093/pcp/pcr156
- Yamashita, Y., Ota, M., Inoue, Y., Hasebe, Y., Okamoto, M., Inukai, T., et al. (2016). Chemical promotion of endogenous amounts of ABA in *Arabidopsis thaliana* by a natural product, theobroxide. *Plant Cell Physiol.* 57, 986–999. doi: 10.1093/pcp/pcw037
- Yang, L., Stulen, I., and De Kok, L. J. (2006). Impact of sulfate nutrition on the utilization of atmospheric SO₂ as sulfur source for Chinese cabbage. *J. Plant Nutr. Soil Sci.* 169, 529–534. doi: 10.1002/jpln.200520574
- Zhang, B., Pasini, R., Dan, H., Joshi, N., Zhao, Y., Leustek, T., et al. (2014). Aberrant gene expression in the *Arabidopsis* SULTR1;2 mutants suggests a possible regulatory role for this sulfate transporter in response to sulfur nutrient status. *Plant J.* 77, 185–197. doi: 10.1111/tpj.12376
- Zhang, J., Sun, X., Zhang, Z., Ni, Y., Zhang, Q., Liang, X., et al. (2011). Metabolite profiling of *Arabidopsis* seedlings in response to exogenous sinalbin and sulfur deficiency. *Phytochemistry* 72, 1767–1778. doi: 10.1016/j.phytochem.2011.06.002
- Zhao, Z., Zhang, W., Stanley, B. A., and Assmann, S. M. (2008). Functional proteomics of *Arabidopsis thaliana* guard cells uncovers new stomatal signaling pathways. *Plant Cell* 20, 3210–3226. doi: 10.1105/tpc.108.063263
- Zhou, C., Tokuhisa, J. G., Bevan, D. R., and Esen, A. (2012). Properties of beta-thioglucoside hydrolases (TGG1 and TGG2) from leaves of *Arabidopsis thaliana*. *Plant Sci.* 191–192, 82–92. doi: 10.1016/j.plantsci.2012.02.004

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Sugiyama and Hirai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Molecular Origins of Functional Diversity in Benzyloquinoline Alkaloid Methyltransferases

Jeremy S. Morris and Peter J. Facchini*

Department of Biological Sciences, University of Calgary, Calgary, AB, Canada

OPEN ACCESS

Edited by:

Kazuki Saito,
RIKEN Center for Sustainable
Resource Science (CSRS), Japan

Reviewed by:

Joerg Ziegler,
Leibniz-Institut für Pflanzenbiochemie
(IPB), Germany
Dietrich Ober,
University of Kiel, Germany

*Correspondence:

Peter J. Facchini
pfacchin@ucalgary.ca

Specialty section:

This article was submitted to
Plant Metabolism and
Chemodiversity,
a section of the journal
Frontiers in Plant Science

Received: 26 April 2019

Accepted: 30 July 2019

Published: 30 August 2019

Citation:

Morris JS and Facchini PJ (2019)
Molecular Origins of Functional
Diversity in Benzyloquinoline
Alkaloid Methyltransferases.
Front. Plant Sci. 10:1058.
doi: 10.3389/fpls.2019.01058

O- and N-methylations are ubiquitous and recurring features in the biosynthesis of many specialized metabolites. Accordingly, the methyltransferase (MT) enzymes catalyzing these modifications are directly responsible for a substantial fraction of the vast chemodiversity observed in plants. Enabled by DNA sequencing and synthesizing technologies, recent studies have revealed and experimentally validated the trajectories of molecular evolution through which MTs, such as those biosynthesizing caffeine, emerge and shape plant chemistry. Despite these advances, the evolutionary origins of many other alkaloid MTs are still unclear. Focusing on benzyloquinoline alkaloid (BIA)-producing plants such as opium poppy, we review the functional breadth of BIA N- and O-MT enzymes and their relationship with the chemical diversity of their host species. Drawing on recent structural studies, we discuss newfound insight regarding the molecular determinants of BIA MT function and highlight key hypotheses to be tested. We explore what is known and suspected concerning the evolutionary histories of BIA MTs and show that substantial advances in this domain are within reach. This new knowledge is expected to greatly enhance our conceptual understanding of the evolutionary origins of specialized metabolism.

Keywords: benzyloquinoline, alkaloid, methyltransferase, specialized metabolism, molecular evolution

INTRODUCTION

The incredible diversity of plant metabolism has been a topic of fascination for centuries, and yet our appreciation for its scope continues to grow. A recent study examined more than a hundred thousand metabolite–plant species relationships and concluded that each species contains, on average, 4.7 unique metabolites, which sum to an estimate of more than 1 million distinct metabolites across the kingdom (Afendi et al., 2012). An earlier analysis of a smaller database calculated the existence of only 1.6 unique metabolites per plant species, suggesting that our estimates of the breadth of plant metabolism will continue to grow as we collect more data (Shinbo et al., 2006). These numbers are in line with previous estimates ranging from 200,000 to 1,000,000 plant metabolites in total (Saito and Matsuda, 2009).

Alkaloids, broadly defined as low-molecular weight heterocyclic nitrogenous compounds, are thought to occur in roughly 20% of plant species (Ziegler and Facchini, 2008). At least 12,000 unique molecules of this class are known, which can be classified as either protoalkaloids (e.g., mescaline, ephedrine), where the nitrogen is not cyclic, pseudoalkaloids (e.g., steroidal and diterpene alkaloids, caffeine), where the primary biosynthetic origin is not an amino acid, and “true” alkaloids where most of the molecule, including the heterocyclic nitrogen, is derived from an amino acid precursor

(Hegnauer, 1988; Waterman, 1998). This latter group is most diverse and includes biosynthetic end products derived from phenylalanine, tyrosine, tryptophan, ornithine, arginine, lysine, histidine, and anthranilic acid.

Although the foundation of plant alkaloid chemical diversity begins with the combination and rearrangement of the aforementioned building blocks, each basic carbon skeleton can give rise to a great number of “decorated” variants with various functional group substitutions that alter the molecule’s biochemical characteristics. For example, *N*-methylation of xanthine/xanthosine during biosynthesis of caffeine allows for the production of up to seven differentially methylated products (Huang et al., 2016). Similarly, the potential for two *N*-methyl and four *O*-methyl groups on simple benzylisoquinoline alkaloids (BIAs) such as norlaudanosoline makes up to 30 distinct molecules possible. The addition of methyl groups to an alkaloid molecule can have important consequences regarding its chemical properties and thus shift biological activity. Methylation can invert the polarity of an electronegative moiety, shift the molecule’s stereoelectronic profile, increase overall hydrophobicity, increase steric bulk, and promote or prevent certain conformations of the molecule (Wessjohann et al., 2014). In the extreme case, methylation of a tertiary amine results in a quaternary ammonium cation, which is substantially more hydrophilic and lipophobic. For example, *O*-methylation of the monoterpene indole alkaloid noribogaine results in a much less polar compound (ibogaine), which is readily sequestered to lipophilic compartments of the mammalian brain (Zubaran, 2000). The *O*-methylated molecule displays differential binding to neurotransmitter receptors versus the parent compound, resulting in substantially more toxicity as measured by the dose required to induce tremors and cerebellar damage. Similarly, the *O*-methylated BIA thebaine is much more of a stimulant and much less of an effective painkiller than morphine, which is fully *O*-demethylated (Navarro and Elliott, 1971).

Plant Alkaloid *O*-Methyltransferases

Underlying the massive number of differentially methylated plant alkaloids is a large and heterogeneous group of methyltransferase (MT) enzymes thought to be specialized for various substrates (Figure 1). Several *O*-methyltransferase (OMT) enzymes, which participate in the terminal steps of monoterpene indole alkaloid biosynthesis, have been identified and cloned, including an OMT leading to the production of vindoline in *Catharanthus roseus* (Levac et al., 2008), an OMT producing ibogaine in *Tabernanthe iboga* (Farrow et al., 2018) and a 10-hydroxycamptothecin OMT from *Camptotheca acuminata* (Salim et al., 2018). Three OMTs contributing to the biosynthesis of monoterpene isoquinoline alkaloids such as emetine have been cloned from *Psychotria ipecacuanha* (Nomura and Kutchan, 2010). Studies on the biosynthesis of Amaryllidaceae alkaloids such as galanthamine in *Narcissus* spp. and *Lycoris aurea* allowed for the isolation of two norbelladine 4′ OMT enzymes (Kilgore et al., 2014; Sun et al., 2018). Although not traditionally included in most lists of alkaloids due to an unclear biosynthetic origin, the volatile heterocyclic nitrogenous methoxypyrazines, which contribute to

the flavor profile of grapes, also require *O*-methylation in their biosynthesis. To date, four *Vitis vinifera* OMTs implicated in this pathway have been cloned (Dunlevy et al., 2010; Dunlevy et al., 2013; Guillaumie et al., 2013). Quite a few OMTs implicated in benzylisoquinoline alkaloid biosynthesis have been cloned, and these will be reviewed in a dedicated section below.

Plant Alkaloid *N*-Methyltransferases

Particularly well studied are *N*-methyltransferases (NMTs), which catalyze the terminal biosynthetic steps producing xanthine alkaloids (e.g., caffeine) in various plants. Cloned representatives include caffeine synthase, theobromine synthase and 7-methylxanthosine synthase from *Coffea arabica* (Mizuno et al., 2003a; Mizuno et al., 2003b) as well as homologs in *Camellia*, *Theobroma*, *Paullinia*, and *Citrus* (Kato et al., 2000; Yoneyama et al., 2006; Schimpl et al., 2014; Huang et al., 2016). Cloned NMTs contributing to the biosynthesis of monoterpene indole alkaloids include one from *C. roseus* leading to the production of vindoline (Liscombe et al., 2010) and two related picrinine NMTs from *Apocynaceae* species (Levac et al., 2016). NMTs from less intensively studied pseudoalkaloid pathways have also been cloned, such as those implicated in the biosynthesis of gramine (Larsson et al., 2006) and ephedrine in *Hordeum vulgare* and *Ephedra sinica*, respectively (Morris et al., 2018). The large number of BIA NMTs that have been characterized at the molecular level will be discussed further in the body of this work.

Although most known alkaloid MTs contribute to the final stages of biosynthesis, this is not a firm rule. Putrescine NMT (PMT), which synthesizes *N*-methylputrescine, catalyzes the first step in pathways leading to several alkaloid classes including the pyridines (e.g., nicotine), tropane alkaloids (e.g., scopolamine), or calystegines in various plant species (Blastoff et al., 2009). A larger number of PMTs have been cloned, including those from *Nicotiana tabacum*, *Solanum tuberosum*, and various other Solanaceae and Convolvulaceae species (Hibi, 1994; Stenzel et al., 2006; Teuber et al., 2007; Junker et al., 2013). Similarly, an anthranilate NMT diverts metabolic flux away from tryptophan biosynthesis into the acridone alkaloid biosynthetic pathway and has been cloned from *Ruta graveolens* (Rohde et al., 2007).

Despite the ever-growing list of characterized and cloned MTs, there remain many biosynthetic pathways to explore. For example, *O*- and *N*-methylation of phenethylamine alkaloids in many lineages including the Cactaceae (e.g., mescaline), tryptamine alkaloids in *Acacia*, *Citrus*, *Phalaris*, and others (e.g., *N,N*-dimethyltryptamine), quinolizidine alkaloids in Fabaceae (e.g., *N*-methylcytisine), and the diverse Amaryllidaceae alkaloids remains understudied (Smith, 1977a; Smith, 1977b; Wink, 1984; Jin and Xu, 2013).

Caffeine Biosynthesis as a Model of Research Potential

The wealth of cloned alkaloid MTs has proven to be a fertile area in which to examine the relationships between enzyme function and plant biochemistry. Aside from the characterization of natural variants and concomitant identification of sequence–function correlations, modern structural biology and DNA manipulation

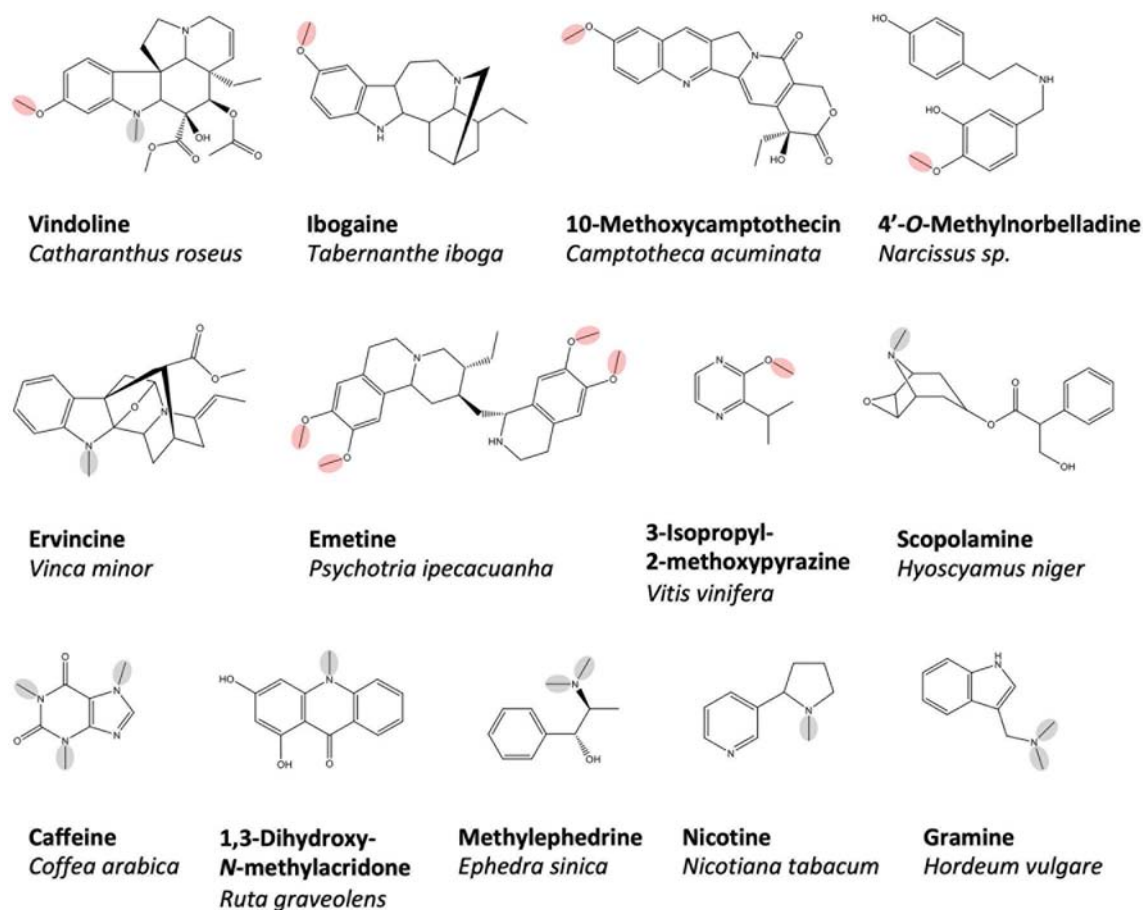


FIGURE 1 | A selection of plant alkaloids biosynthesized by O- and N-methyltransferases, which have been cloned and functionally characterized *in vitro*. O- and N-methyl groups reportedly installed by the enzymes are circled in red and green, respectively. Representative benzyloisoquinoline alkaloids produced are shown in **Figure 2**.

methods have allowed experimental approaches to directly probe the features controlling enzyme properties. In select cases, the molecular evolutionary trajectories, which resulted in extant enzyme, and their specific properties have also been elucidated, providing some insight into the origins of specialized biochemical pathways and the exceptional chemodiversity of plants.

Perhaps the best example of the research sequence described above relates to the biosynthesis of caffeine and other xanthine alkaloids. Building on a long history of research using radiolabeled tracers to elucidate the pathway, workers eventually showed unequivocally that caffeine is synthesized from xanthosine *via* a series of N-methylation reactions (Suzuki and Takahashi, 1975; Suzuki and Takahashi, 1976; Ashihara et al., 1996; Kato et al., 1996; Mösl Waldhauser et al., 1997a; Mösl Waldhauser et al., 1997b; Kato et al., 1999). Shortly thereafter, cDNAs encoding these enzymes were cloned from *Camellia sinensis* (Kato et al., 2000) and *Coffea arabica* (Ogawa et al., 2001; Uefuji, 2003; Mizuno et al., 2003a, Mizuno et al., 2003b) and later from *Paullinia cupana* (Schimpl et al., 2014), *Theobroma cacao* (Yoneyama et al., 2006), and *Citrus sinensis* (Huang et al., 2016). Examination of their coding sequences showed that these xanthine NMTs belong to the

SABATH (salicylic acid, benzoic acid, theobromine) family of methyltransferases, which typically methylate oxygen atoms, suggesting a relatively recent change of function in xanthine alkaloid-producing species. Intriguingly, greater sequence similarity between the functionally distinct MTs within one species (e.g., 80% identity between CaXMT, CaMXMT, and CaDXMT from *C. arabica*) compared to those of analogous function in other plants (e.g., less than 40% identity between CaDXMT and TCS1 from *C. sinensis*) lead to the hypothesis that xanthine MTs in different lineages have parallel and convergent evolutionary histories.

With coding sequences available to produce heterologous proteins, the molecular determinants of function were studied (Yoneyama et al., 2006; McCarthy and McCarthy, 2007). Active-site features facilitating binding of xanthine molecules, in general (e.g., hydrophobic pocket), and individual residues implicated in substrate specificity (e.g., Ser316 and Tyr356 in CaXMT hydrogen bonding with xanthine but not methylxanthine) were proposed based on comparative analysis of crystal structures in complex with various substrates. Next, site-directed mutagenesis allowed for experimental validation of these hypotheses (Yoneyama et al., 2006; Huang et al., 2016; Jin et al., 2016).

More recently, whole genome sequencing of *Coffea*, *Camellia*, and *Theobroma* species began to unveil the genetic mechanisms leading to the evolution of caffeine biosynthesis (Argout et al., 2011; Denoeud et al., 2014; Xia et al., 2017; Wei et al., 2018). Taken together, the studies strongly suggested the occurrence of multiple independent and convergent evolutionary trajectories, in which gene duplication and functional divergence lead to caffeine biosynthesis. However, these results did not yet explain how or why this biochemical feature arose so readily in distantly related plants.

Clarification of one evolutionary trajectory leading to caffeine biosynthesis in *Citrus* was recently provided by a paleomolecular biology method known as ancestral enzyme reconstruction (Thornton, 2004; Huang et al., 2016). Barkmann and colleagues showed that an ancestral SABATH enzyme was likely exapted to catalyze *N*-methylation of caffeine pathway intermediates and that, following its duplication, a single amino acid substitution in each of the descendant enzymes was sufficient to create a fully functional caffeine biosynthetic pathway. Their results concerning the unusually short mutational distance between SABATH OMT and xanthine NMT functions showed how evolutionary innovation was able to repeatedly converge on caffeine biosynthesis in multiple plant orders.

The extent of our knowledge surrounding the evolution of caffeine biosynthesis showcases what can be achieved with modern research tools and paradigms. In the following section, we outline the current state of knowledge regarding *O*- and *N*-methyltransferases involved in BIA biosynthesis and their contribution to host plant chemodiversity. Drawing on structural and functional studies, we outline what is known regarding the molecular determinants of their differing activities and explore what is suspected regarding their evolution. We show that the field is ripe for substantial advances paralleling those obtained with respect to caffeine biosynthesis and suggest key hypotheses and experiments by which they may be tested.

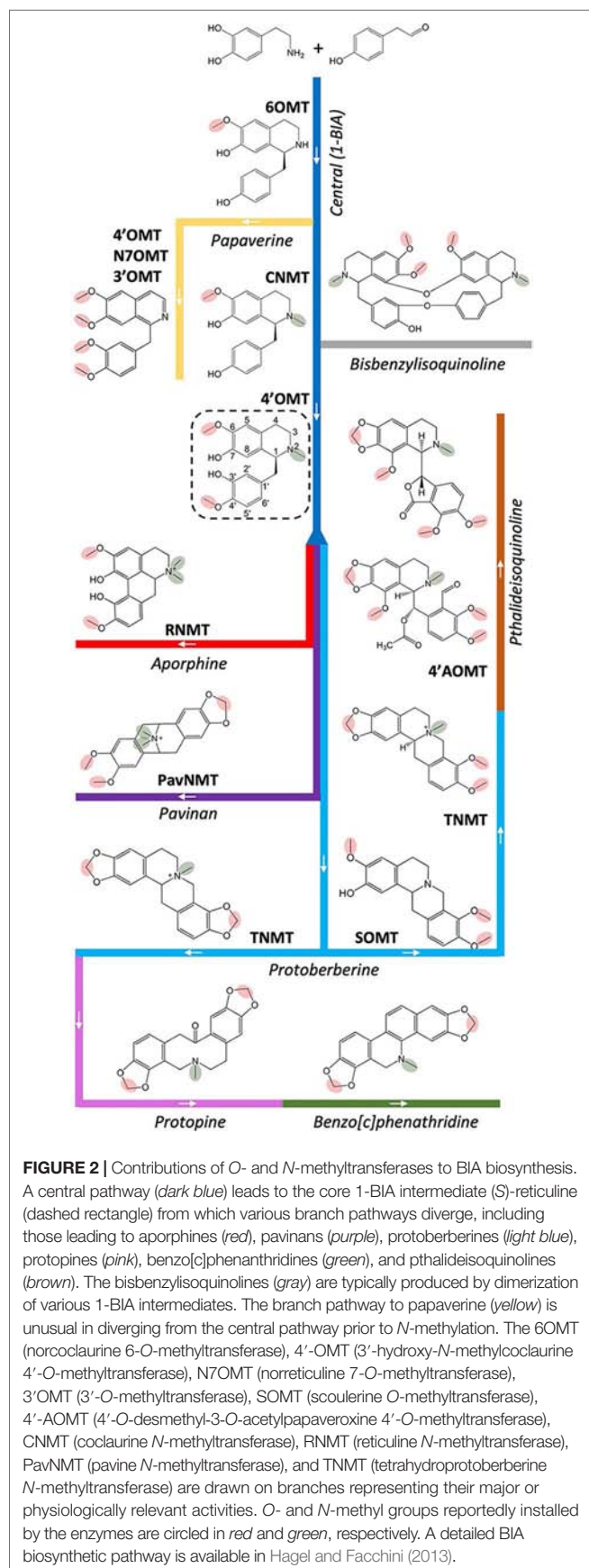
CONTRIBUTION OF METHYLTRANSFERASES TO BENZYLISOQUINOLINE ALKALOID BIOSYNTHESIS

BIAs have been studied for centuries (Hagel and Facchini, 2013 and references therein), and much of their biosynthesis has been revealed, albeit only in a handful of model systems, which can only approximate the biosynthetic diversity in the thousand or more BIA-producing plant species (Shulgin and Perry, 2002). These discoveries and their historical context have been extensively reviewed elsewhere and will only be summarized here in order to highlight the involvement of methyltransferases. A tremendous number of BIA MTs have been characterized in plant extracts, and although these have contributed greatly to our understanding of BIA biosynthesis, we will focus herein on those that have been cloned and studied at the molecular level. The nine Ranunculales species from which BIA MTs have been characterized at this level are *Papaver somniferum*,

P. bracteatum, *Glaucium flavum*, *Thalictrum flavum*, *Coptis japonica*, *C. chinensis*, *C. teeta*, *Dactylicapnos scandens*, and *Eschscholzia californica*. Literature reports concerning the isolation and characterization of each of the MTs discussed are cited in **Supplementary Table 1**.

In *P. somniferum*, the most studied BIA model organism, a central pathway for BIA biosynthesis begins with a condensation of two tyrosine derivatives (dopamine and 4-hydroxyphenylacetaldehyde) to form (S)-norcoclaurine (**Figure 2**). Acting on this base skeleton, norcoclaurine-6-*O*-methyltransferase (6OMT) transfers a methyl group onto one of the isoquinoline moiety hydroxyl groups to yield (S)-coclaurine. Given the core role of this enzymes, it is unsurprising that nine cognate cDNAs from six species have been isolated and shown to encode this activity (**Table 1**; **Supplementary Table 1**) (Morishige et al., 2000; Ounarooun et al., 2003; Facchini and Park, 2003; Samanani, 2005; Tamura et al., 2006; Desgagné-Penix and Facchini, 2012; Chang et al., 2015; Robin et al., 2016; He et al., 2018). Interestingly, several species (e.g., *G. flavum*, *C. chinensis*) seem to express multiple distinct transcripts encoding enzymes with this activity despite sharing only 45–55% amino acid identity (**Supplementary Figure 1**). Downstream in the central pathway, a first *N*-methyl group is installed by coclaurine *N*-methyltransferase (CNMT), which has been cloned from four BIA-producing species (Choi et al., 2002; Facchini and Park, 2003; Samanani, 2005; Minami et al., 2008; Liscombe et al., 2009; Desgagné-Penix and Facchini, 2012; Hagel et al., 2015). Notably, four transcripts encoding enzymes with CNMT activity (59–66% AA identity; **Supplementary Figure 2**) have been cloned from *G. flavum*; however, their individual contributions to biosynthesis in the host plant remain to be assessed. The final step in the central pathway is catalyzed by 3'-hydroxy-*N*-methylcoclaurine 4'-*O*-methyltransferase (4'OMT) and yields the triple-methylated central intermediate (S)-reticuline. Corresponding transcripts have been cloned from three species (Morishige et al., 2000; Facchini and Park, 2003; Ounarooun et al., 2003; Ziegler et al., 2005; Desgagné-Penix and Facchini, 2012; Chang et al., 2015). Widespread occurrence of the above three MTs in BIA-producing species is consistent with the current model in which all end product alkaloids derive from reticuline or, less commonly, from upstream central pathway intermediates (Hagel et al., 2015).

The 4'-*O*-methylation also contributes to papaverine biosynthesis via a branch that diverges from the central pathway prior to the action of CNMT and, instead, passes through (S)-norreticuline. Next, the exceptionally substrate-specific norreticuline 7-*O*-methyltransferase (N7OMT) installs a third *O*-methyl to yield (S)-norlaudanine. The single N7OMT cloned to date occurs in *P. somniferum*, which is consistent with the somewhat-restricted taxonomic distribution of papaverine (Shulgin and Perry, 2002; Pienkny et al., 2009; Desgagné-Penix and Facchini, 2012). Although the preponderance of evidence at this time supports an *N*-desmethylated biosynthetic scheme for papaverine, it was also proposed that biosynthesis might pass through (S)-reticuline. Indeed, a reticuline 7-*O*-methyltransferase (7OMT) has also been cloned from *P. somniferum*, along with eight additional transcripts encoding 7OMT-like enzymes from five species (Ounarooun et al., 2003; Fujii et al., 2007; Dang and



Facchini, 2012; Desgagné-Penix and Facchini, 2012; Chang et al., 2015; Purwanto et al., 2017; He et al., 2018). The 3'-O-methylation, which completes the series of four methylations required to yield papaverine, may be catalyzed in part via *P. somniferum* SOMT1 despite the fact that its *in vitro* activity substantially favors other substrates (Dang and Facchini, 2012). An additional five transcripts encoding enzymes with comparable 3'OMT activities have been cloned from three species (Fujii et al., 2007; Chang et al., 2015; Purwanto et al., 2017). Aside from papaverine, many 3'- and 7-O-methylated simple BIAs (e.g., laudanin, laudanin) and potential derivatives (e.g., jatrorrhizine, tetrahydropalmatine) are known; however, the involvement of any given OMT in their biosynthesis remains speculative due to a lack of *in planta* experimental evidence.

Starting from (S)-reticuline, a short branch yields the taxonomically widespread aporphine alkaloid (S)-magnoflorine (e.g., Ranunculales, Laurales, Magnoliales, Sapindales, Piperiales, etc.) (Shulgin and Perry, 2002). Acting on either reticuline or its aporphine derivative (corytuberine), *P. somniferum* reticuline N-methyltransferase (RNMT, named as such to reflect *in vitro* substrate preference; Table 1; Supplementary Table 1) installs a second methyl group resulting in a quaternary nitrogen atom. To date, three additional transcripts encoding RNMT-like enzymes have been reported in two other BIA-producing plant species, both of which are also known to accumulate (S)-magnoflorine and related quaternary alkaloids (Shulgin and Perry, 2002; Liscombe et al., 2009; Hagel et al., 2015; Torres et al., 2016). Of particular note is the enzyme cloned from *T. flavum* and named pavine N-methyltransferase (TfPavNMT). Whereas the enzyme efficiently catalyzes the signature RNMT-like activity and might contribute to quaternary aporphine biosynthesis, it also uniquely accepts pavinan alkaloids and is thought to participate in the biosynthesis of N-methylescholzidine, an uncommon BIA, which accumulates in *Thalictrum* spp (Liscombe et al., 2009). In *D. scandens*, a corytuberine 7-O-methyltransferase (DsC7OMT) was identified, and preliminary analysis linked it to the biosynthesis of isocorydine (He et al., 2017).

Several longer branch pathways employing MT reactions also diverge from (S)-reticuline to produce protoberberines (e.g., stylopine), protopines (e.g., protopine), benzo[c]phenanthridines (e.g., sanguinarine), and phthalideisoquinolines (e.g., noscapine) (Figure 2). Acting on the first protoberberine intermediate, scoulerine 9-O-methyltransferase (SOMT) installs a third O-methyl group to yield tetrahydrocolumbamine. In agreement with the relatively common occurrence of protoberberines in BIA-producing plants, known enzymes with SOMT activity are encoded by 12 transcripts isolated from five different species (Takeshita et al., 1995; Morishige et al., 2000; Fujii et al., 2007; Dang and Facchini, 2012; Chang et al., 2015; Purwanto et al., 2017; He et al., 2018). A functionally similar enzyme, which preferentially targets the 2-hydroxyl of quaternary protoberberine columbamine, was cloned only from *C. japonica* (CjCoOMT) (Morishige et al., 2002). En route to noscapine biosynthesis, tetrahydropprotoberberine NMT (TNMT) transfers a methyl group onto a bicyclic nitrogen atom, yielding the quaternary product N-methylcanadine. In addition to the canonical representative isolated from *P. somniferum*, six

TABLE 1 | O- and N-methyltransferases reportedly able to catalyze the various major BIA biosynthetic activities *in vitro*. Details of the activities and corresponding citations are provided in **Supplementary Table 1**.

6OMTs	4'OMTs	N7OMT	7OMTs	3'OMTs	SOMTs	4'AOMTs	CNMTs	RNMTs	TNMTs
GfiOMT1	GfiOMT1	PsN7OMT	GfiOMT1	PsSOMT1	GfiOMT1	PsSOMT2: SOMT3	GfiNMT1	PsRNMT	GfiNMT2
GfiOMT2	GfiOMT2		GfiOMT6	GfiOMT1	GfiOMT2	PsSOMT2: 6OMT	GfiNMT4	GfiNMT4	GfiNMT3
Ps6OMT	Ps4'OMT2		PsSOMT1	GfiOMT2	GfiOMT6		GfiNMT5	GfiNMT5	PsTNMT
Ps7OMT	Ps7OMT		Ps7OMT	GfiOMT6	GfiOMT7		GfiNMT6	TfPavNMT	EcTNMT
Cj6OMT	Cj4'OMT		Cc6OMT1	Ec7OMT	PsSOMT1		PsCNMT	EsPaNMT*	PbTNMT
Cc6OMT1			Cc6OMT2	EcS/ROMT	PsSOMT2		CjCNMT		TfPavNMT
Cc6OMT2			Ct6/7OMT		PsSOMT3		TfCNMT		TfCNMT
Ct6/7OMT			Ec7OMT		CjSOMT		EsPaNMT*		EsPaNMT*
Tf6OMT			EcS/ROMT		CjCoOMT				
					CtSOMT				
					Ec7OMT				
					EcS/ROMT				

6OMT, 6-O-methyltransferase; 4'OMT, 4'-O-methyltransferase; N7OMT, norreticuline 7-O-methyltransferase; 7OMT, 7-O-methyltransferase; 3'OMT, 3'-O-methyltransferase; SOMT, scoulerine-O-methyltransferase; 4'AOMT, 4'-O-desmethyl-3-O-acetylpapaverine 4'-O-methyltransferase; CNMT, coclaurine N-methyltransferase; RNMT, reticuline N-methyltransferase; TNMT, tetrahydroprotoberberine N-methyltransferase. *Ephedra sinica* phenylalkylamine N-methyltransferase (EsPaNMT), denoted with an asterisk, has promiscuous activity with various BIAs but does not participate in their biosynthesis in planta. *Dactylicapnos scandens* coryuberine 7-O-methyltransferase (DsC7OMT) and *Eschscholzia californica* 10-hydroxysanguinarine O-methyltransferase (EcG11OMT) have unique substrate ranges and are not included in the table.

other TNMT-like enzymes have been cloned from four other species (Samanani, 2005; Liscombe and Facchini, 2007; Liscombe et al., 2009; Hagel et al., 2015; Torres et al., 2016). Intriguingly, the final O-methylation required to produce noscapine was recently shown to involve a heterodimer composed of PsSOMT2 and either PsSOMT3 or Ps6OMT (Li and Smolke, 2016; Park et al., 2018). No equivalent enzymes are presently known in other species, which is consistent with the lack of reports of noscapine in BIA producing species outside Papaveraceae (Shulgin and Perry, 2002). Aside from the role outlined above, TNMTs also participate in a separate branch pathway leading first to protopines and then to benzo[c]phenanthridines *via* synthesis of (S)-*cis*-N-methylstylopine. This product accumulates to a substantial degree in *T. flavum*, which is inconsistent with the relatively modest stylopine NMT activity reported for TfPavNMT (Liscombe et al., 2009). Thus, it seems likely that one of the additional NMT transcripts recently identified in that species' transcriptome encodes an enzyme more dedicated to protoberberine substrates (Hagel et al., 2015). Acting further downstream in the benzo[c]phenanthridine branch pathway, a functionally unique OMT from *E. californica* (EcG11OMT) apparently targets 10-hydroxysanguinarine (Purwanto, 2017).

Aside from the above, a number of transcripts have been cloned and found to encode enzymes with high homology to BIA MTs but with no discernible activity *in vitro* (Liscombe and Facchini, 2007; Chang et al., 2015; Morris and Facchini, 2016; Purwanto et al., 2017; He et al., 2018). Although these may, in fact, be inactive with respect to BIAs, it remains possible that they have not been assayed under appropriate conditions or with proper substrates. These mysterious transcripts include at least one NMT each from *P. somniferum* and *Arabidopsis*, as well as a large number of OMTs from *G. flavum*, *P. somniferum*, *E. californica*, and *Coptis* spp. In the latter group, substantial homology to other types of plant OMTs makes identification of those targeting BIAs quite challenging.

As revealed in **Table 1**, the majority of BIA MTs are known to catalyze many additional reactions beyond those prototypical conversions represented by their names and position on orderly, linear biosynthetic pathways as traditionally drawn (**Figure 2**). Nevertheless, the targeted nature of *in vitro* biochemical characterization means that all reports necessarily underestimate the catalytic range of BIA MTs. The lack of specificity reported for BIA MTs occurs at two levels: substrate promiscuity, wherein the enzyme can methylate a number of different molecules, and product promiscuity, wherein a single substrate is methylated one or more times at various positions to yield different products (O'Brien and Herschlag, 1999; Hult and Berglund, 2007). On the other hand, BIA MTs do not show catalytic promiscuity, which is the ability to carry out distinct types of chemical transformations. Thus far, BIA OMTs and NMT have only been shown to catalyze O- and N-methylation, respectively, unlike certain SABATH MTs that target both O and N atoms (Huang et al., 2016). As shown for PsRNMT with respect to magnoflorine biosynthesis, the most substantial activity of an enzyme *in vitro* may not correlate with its function *in planta* (Morris and Facchini, 2016). This widespread promiscuity has led to an appreciation for the existence of multidimensional "metabolic grids", which diversify the potential routes by which a plant may make any given end product. Experimental evidence for major, minor, or even "silent" routes in BIA biosynthesis has been given by gene knockdown experiments in whole plants and cell cultures (Fujii et al., 2007; Desgagné-Penix and Facchini, 2012).

MOLECULAR AND STRUCTURAL DETERMINANTS OF FUNCTION

O-Methyltransferases

To date, structures have been reported for *T. flavum* norcoclaurine 6OMT (Tf6OMT; PDB 5ICE) and *P. somniferum* scoulerine

O-methyltransferase 1 (PsSOMT1; PDB 6I6K) (Robin et al., 2016; Cabry et al., 2019). While these structures have allowed for the generation of compelling hypotheses concerning substrate binding and catalysis, relatively little experimental work (e.g., site-directed mutagenesis) is presently available in support of their validity. The overall structures, each composed of an N-terminal dimerization domain linked to a C-terminal substrate binding domain, are consistent with those previously reported for SAM-dependent OMTs in plants such as *Medicago sativa* Caffeic acid OMT (MsCOMT; PDB 1KYZ), *M. trunculata* isoflavonoid OMT (MsIOMT; PDB 1FP2), and *M. sativa* chalcone OMT (MsChOMT; PDB 1FP1) (Zubieta et al., 2001; Zubieta, 2002).

Dimerization

BIA OMTs form dimers in solution as well as in all obtained crystal structures. Dimerization occurs *via* a substantial (150 of 350 residues in Tf6OMT) domain composed primarily of intertwined helices (**Figure 3**). Most of these helices interact with those of the other monomer, resulting in burial of approximately 22% of total protein surface area. Across BIA OMTs, the entire dimerization domain shows relatively modest conservation. Nevertheless, two leucine residues (Leu28, Leu73) show perfect identity, and an additional four hydrophobic residues (Leu34, Ile40, Leu52, Leu66) show strong conservation (**Supplementary Figure 3**). However, their contributions to the dimer interface do not appear to be substantial, suggesting that they may be more important in maintaining secondary and tertiary structures of each monomer (PDBePISA) (Krissinel and Henrick, 2007). Surprisingly, gel filtration analysis of CjSOMT leads authors to report the existence of a trimer in solution (Morishige et al., 2000). Examination of the CjSOMT sequence does reveal a 20-amino acid extension of the N-terminus relative to Tf6OMT. However, it is unclear how this feature could so drastically alter how the monomers associate.

Although heterologously expressed BIA OMTs have generally been studied under the assumption that they form homodimers *in planta*, several studies suggest that dimerization of genetically distinct monomers (i.e., heterodimerization) may play a crucial role in alkaloid biosynthesis (Frick and Kutchan, 1999; Li and Smolke, 2016; Park et al., 2018). Recently, the missing methylation step in *P. somniferum* noscapine biosynthesis was shown to involve a heterodimer of PsSOMT2 and either PsSOMT3 or Ps6OMT (Park et al., 2018). The discovery of an OMT heterodimer that catalyzes a reaction not performed by either homodimer reveals an unexpected diversification strategy by which additional OMT activities can exist without the genomic and metabolic burden of maintaining and expressing dedicated OMT genes. Although the extent of heterodimerization in plant specialized metabolism remains to be established, it is intriguing to consider that many more OMT activities might be linked to heterodimers in the future.

Several helices of the dimerization domain also contribute to the active site, which is found at the interface between the first monomer's C-terminal domain, two helices of its N-terminal domain, and two helices belonging to the N-terminal domain of the second monomer (**Figure 3**). Together, these four helices form a hydrophobic "back wall" at the active site, but no direct interactions with the substrate have been reported. The lack of

clear interactions is surprising given several pieces of evidence showing that the N-terminal domains of both monomers make contributions to substrate selectivity. For example, a chimeric enzyme fusing the Cj6OMT dimerization domain to the C-terminal domain of Cj4'OMT displayed substrate- and regio-specificity most similar to Cj6OMT (Morishige et al., 2010). Further shuffling of the two polypeptide sequences suggested that the determinants of function exist between the 34th and 125th amino acids in Cj6OMT, which is a region that includes the two helices mentioned above, which contribute to a monomer's own active site, but excludes the two helices, which contribute to the second monomer's active site. Other lines of evidence suggest that the identity of the second monomer, potentially mediated by the aforementioned two helices, alters substrate specificity in the first monomer's active site. It was shown that although the PsSOMT2 monomer contains the catalytic machinery necessary for turnover in the PsSOMT2:PsSOMT3 heterodimer, substitution of PsSOMT3 with either PsSOMT2 or PsN7OMT abolished activity (Park et al., 2018). Structural and functional mutagenesis studies should help clarify the origin of these indirect effects.

SAM Binding

Most of the BIA OMT polypeptide (200 of 350 residues in Tf6OMT) forms a C-terminal domain composed of alternating alpha helices and beta sheets which together create a Rossmann fold classically associated with nucleotide binding (**Figure 3**). SAM binding occurs *via* a series of motifs, which are highly conserved with other plant SAM-dependent MTs (Kozbial and Mushegian, 2005; Gana et al., 2013). The residues within these motifs, which directly interact with SAM, are almost perfectly conserved in known BIA OMTs. However, three SOMTs display conservative (e.g., Asp to Glu) or semi-conservative (e.g., Asp to Gln) substitutions at two positions (**Supplementary Figure 3**). Unfortunately, neither of these substitutions occurs in the recently crystallized PsSOMT1, leaving the question of how these might alter OMT-SAM interactions unresolved. Comparison of apo-enzyme and enzyme-substrate complexes further reveals that the residue equivalent to Thr170 in Tf6OMT (Ser211 in PsSOMT1) interacts with the co-substrate upon binding and contributes to a substantial conformational change (16° hinge movement) of the enzyme likely to be crucial for catalysis (Robin et al., 2016). Intriguingly, binding of SAH was sufficient to induce the closing movement in Tf6OMT, whereas the equivalent movement was only observed after binding of both SAH and the BIA substrate in PsSOMT1 (Cabry et al., 2019). Despite this minor difference, the available evidence suggests that SAM binding occurs in a very similar manner for all BIA OMTs.

BIA Binding

Binding of the alkaloid substrate occurs at a location proximal to the SAM binding site and *via* residues overlapping the aforementioned conserved SAM-binding motifs (**Figure 3**; **Supplementary Figure 3**). Notably, the relative orientation of the bound BIA molecule is flipped between Tf6OMT and PsSOMT1, such that isoquinoline and benzyl moieties or their derivatives in protoberberines (**Supplementary Figure 4**; **Supplementary Figure 5**) generally make reciprocal interactions in either

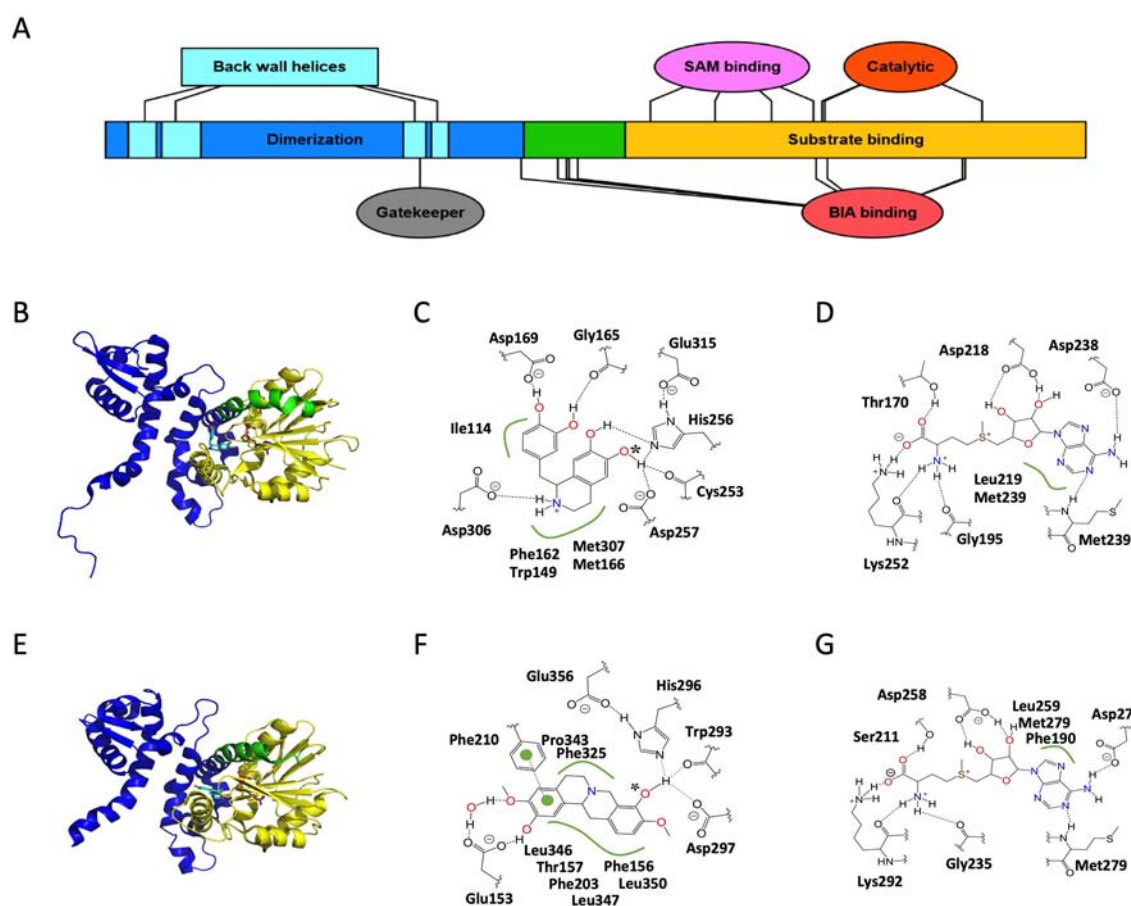


FIGURE 3 | Structure and major active site interactions of representative BIA O-methyltransferases. **(A)** Key features of BIA OMTs are labeled according to their relative location along the *T. flavum* norcoclaurine-6-O-methyltransferase (Tf6OMT) polypeptide with residues denoted by ovals and larger features by rectangles. Crystal structures of **(B)** Tf6OMT with bound SAH and (S)-norlaudanosoline by Robin et al. (2016) (PDB 5ICE), and **(E)** *P. somniferum* scoulerine OMT1 (PsSOMT1) with bound SAH and (S)-scoulerine reported by Cabry et al. (2019) (PDB 6I6K). Domains correspond to those shown in **(A)** and **Supplementary Figure 3**. The alkaloid substrate is shown in cyan, and the cosubstrate SAH is shown in orange. Major active site interactions in **(C, D)** Tf6OMT and **(F, G)** PsSOMT1 were visualized using PoseView and modified where necessary to reflect literature reports. Dashed lines indicate hydrogen bonds, and green lines indicate hydrophobic or aromatic interactions. Dashed lines drawn in gray indicate multiple alternative potential interactions. The oxygen atom accepting the methyl group is identified with an asterisk. **(A)** Drawn using Illustrator for Biological Sequences (Liu et al., 2015).

enzyme (Robin et al., 2016; Cabry et al., 2019). A histidine residue (His256 in Tf6OMT), in which hydrogen bonds with the target hydroxyl group, is perfectly conserved across BIA OMTs. In Tf6OMT, two closely adjacent residues (Asp257 and Cys253) also interact with the target hydroxyl group and likely fine tune its position and reactivity. On the other hand, the equivalent residues in PsSOMT1 (Asp297 and Trp293) were interpreted as forming a channel through which the SAM methyl group is directed toward the acceptor hydroxyl. Although Asp297 is not proposed to be catalytic *per se*, alanine substitution at this position sharply reduced the activity of PsSOMT1 (Cabry et al., 2019). Precise substrate positioning within the active site is also affected by hydrogen bonds formed between non-targeted moieties and residues equivalent to Gly165 (main chain carbonyl) and Asp169 in Tf6OMT. Whereas the former shows strong semi-conservation (i.e., Gly or Ala), the latter is much more variable while maintaining conservation within certain

BIA OMT subtypes (e.g., Glu in 4'OMTs, Asp in 6/7OMTs, His/Phe/Gly in SOMTs). In PsSOMT1, the equivalent residue (Phe210) interacts with a substrate aromatic ring rather than a hydroxyl group.

As seen in the Tf6OMT structure with bound (S)-norlaudanosoline, only a small adjustment in the angle of the substrate is required to place the 7-hydroxyl in an alignment productive for methyl transfer and, in fact, 7-O-methylation activity has not been ruled out for Tf6OMT (Figure 3) (Robin et al., 2016). Conversely, PsSOMT1 has been shown to catalyze both 9- and 2-O-methylation of (S)-scoulerine despite this requiring the substrate to bind in two completely different orientations (Dang and Facchini, 2012). In the absence of crystal structures revealing the details of such alternative binding modes, biophysical modeling supported by mutagenesis studies may be the only practical method by which to understand the determinants of regio-specificity in BIA OMTs.

In Tf6OMT, Asp306 makes a particularly important contribution to alkaloid binding *via* a hydrogen bond with the nitrogen atom (Robin et al., 2016). A close examination of sequence variation among known BIA OMTs at this position reveals a possible explanation for selectivity concerning *N*-methylation status (i.e., un-methylated secondary or mono-methylated tertiary). Whereas PsN7OMT has an aspartic acid residue capable of hydrogen bonding with the secondary nitrogen atom of (*S*)-norreticuline, Ps7OMT and Ec7OMT have leucine and glycine residues instead (**Supplementary Figure 3**). These smaller and uncharged residues might be expected to alleviate steric hindrance, which would occur with the additional methyl group present on tertiary BIAs like (*S*)-reticuline, thus facilitating binding and catalysis. Although the residue equivalent to Asp306 shows strong conservation among BIA OMTs targeting substrates with the simple 1-BIA scaffold (i.e., 6OMT, N7OMT, 4'OMT, etc.), the equivalent is typically hydrophobic (e.g., Leu or Ala) in SOMT-like enzymes. Examination of the PsSOMT1 structures reveals that this substitution pattern likely relates to the “flipped” substrate-binding mode described above. Rather than being positioned near the nitrogen atom, the equivalent residue (Leu346) simply forms a hydrophobic interaction with carbon atoms in one of the adjacent rings.

A number of more generic interactions are also involved in OMT-BIA binding. Two almost perfectly conserved methionine residues (Met166 and Met307 in Tf6OMT) support and position the rings of the isoquinoline moiety *via* sulfur–aromatic interactions (Reid et al., 1985; Robin et al., 2016). The equivalent residues in PsSOMT1 (Met207 and Leu347) instead sandwich the aromatic ring, which derives from the benzyl moiety in 1-BIAs (**Supplementary Figure 3**; **Supplementary Figure 4**) (Cabry et al., 2019). Similarly, two well-conserved aromatic residues (Phe162 and Trp149 in Tf6OMT, Phe190 and Phe203 in PsSOMT1) provide aromatic interactions with the isoquinoline moiety in Tf6OMT but with the benzyl moiety derivative in PsSOMT1. An additional hydrophobic residue (Ile114 in Tf6OMT, Thr157 in PsSOMT1), which helps position the benzyl or isoquinoline moieties, shows much less conservation. Intriguingly, the enzymes, which catalyze *O*-methylation of the benzyl moiety (e.g., Cj4'OMT, Tf4'OMT, Ps4'OMT2) are substituted with a methionine at this position. Given the observation that two methionine residues help position the isoquinoline moiety in Tf6OMT, it is tempting to speculate that an equivalent interaction takes place in 4'OMTs albeit with the BIA molecule positioned such that the isoquinoline and benzyl moieties swap places. This hypothesis is supported by the “flipped” BIA binding pose of PsSOMT1, which reportedly *O*-methylates the benzyl moiety of certain 1-BIAs (Dang and Facchini, 2012). Crystal structures of 4'OMTs will be helpful in testing this hypothesis. More generally, the size and shape of the substrate pocket is thought to control selectivity at a coarse level (e.g., between BIAs, chalcones, or isoflavones). In particular, the bulky Phe156 in PsSOMT1 was recently proposed to act as a “gatekeeper” residue preventing, *via* steric hindrance, binding of substrates with large groups opposite the target hydroxyl (Cabry et al., 2019). In Tf6OMT, a much smaller residue (Thr113) shapes the binding pocket such that the bulkier isoquinoline moiety can be accommodated. This model is an important first step in establishing a unified framework by which to understand and predict substrate specificity of plant OMTs *a priori*.

Catalysis

The catalytic mechanism of BIA OMTs is thought to be conserved with other plant SAM-dependent OMTs (Robin et al., 2016; Cabry et al., 2019). Briefly, a histidine residue (His256 in Tf6OMT, His296 in PsSOMT1) acts as a general base and deprotonates the target hydroxyl group. Subsequently, the newly generated oxyanion carries out a nucleophilic attack on the labile methyl group of SAM, which results in its transfer. The significance of the histidine residue in BIA OMTs was confirmed by targeted mutagenesis experiments, which almost entirely abolished catalytic activity in PsSOMT1 and other *P. somniferum* OMTs (Park et al., 2018; Cabry et al., 2019). An adjacent residue (Asp257 or Asp297), discussed above in the context of substrate binding, can also be thought of as participating in catalysis. In PsSOMT1, substitution of this residue with an alanine yielded an enzyme with roughly 2% activity, leading the workers to conclude that it has an important but non-essential role (Cabry et al., 2019). Structural analysis further implicates a glutamic acid residue (Glu356 in PsSOMT1, Glu315 in Tf6OMT) in hydrogen bonding with the catalytic histidine to promote its necessary basicity. Although the equivalent residue was not discussed with respect to Tf6OMT, examination of their published structures suggests that such an interaction is also present. In fact, perfect conservation of this residue suggests that this aspect of catalysis is maintained in all known BIA OMTs (**Supplementary Figure 3**).

Feedback inhibition is known to be a significant feature affecting BIA OMT activity (Sato et al., 1994; Robin et al., 2016). As shown for Tf6OMT, pathway end products can compete for the active site and thus slow or prevent catalysis. In the case of inhibition by sanguinarine, binding occurs in a position that partially overlaps with that of a productive substrate as described above. Several of the generic interactions (e.g., those with Met166 and Ile114) are preserved, while one of the hydrogen bonding residues (Asp169) interacts with a different O atom. As a result, the planar sanguinarine molecule binds in a position rotated roughly 90° on two axes relative to the productive BIA substrate and forms several new aromatic and hydrogen bonding interactions. Comparison of enzyme–inhibitor interactions to those with a productive substrate can provide rational targets for mutagenesis (i.e., residues interacting with the inhibitor but not the productive substrate), thus potentially allowing for the engineering of feedback-insensitive OMT enzymes, which would be highly useful in biotechnological applications. Unfortunately, the specifics of these interactions are likely to vary significantly from one enzyme–inhibitor pair to the next, thus limiting our ability to generalize from structures already reported in the literature. Although additional crystal structures with bound inhibitors are the gold standard, biophysical modeling approaches (e.g., docking) could also provide some insight on shorter timescales.

N-Methyltransferases

Compared to the OMTs, BIA NMTs have received substantial attention in terms of structure–function investigations despite the much smaller number of functionally characterized representatives. An initial investigation was reported for

T. flavum pavine NMT (PDB 5KOK) and, recently, for *C. japonica* coclaurine NMT (PDB 6GKV) (Torres et al., 2016; Bennett et al., 2018). A third report concerning the tetrahydroprotoberberine NMT from *G. flavum* was accepted for publication during the preparation of this manuscript (PDB 6P3O) (Lang et al., 2019). Taken together, these three studies cover much of the functional range reported to date for BIA NMTs and reveal many of the molecular determinants of function.

The overall structures (Figure 4; Supplementary Figure 6), which include a canonical Rossmann SAM-binding domain as well as a C-terminal substrate-binding domain, are consistent with those reported for other SAM-dependent NMTs including *Plasmodium* phosphoethanolamine NMT (PfPMT; PDB 3UJA) and *Mycobacterium tuberculosis* cyclopropane synthase (MtPcaA; PDB 1KPH) (Huang et al., 2002; Lee et al., 2012). Although we annotate the BIA NMT polypeptide sequences with distinct SAM and BIA substrate-binding regions for simplicity, interacting residues are not strictly found within these regions, and domains become more evident in the tertiary structure. In addition to the typical NMT domains, the BIA NMTs also contain an N-terminal extension composed of three helices, which wrap around the substrate-binding domain and contribute to a homodimerization interface. This N-terminal extension also contributes to positioning a loop and helix proposed to gate the active site.

Dimerization

As with the OMTs, BIA NMTs form dimers both in solution and under crystallization conditions (Torres et al., 2016; Bennett et al., 2018). However, the extent of the dimerization interface is substantially less and corresponds to only 6.5% and 6.9% of surface area in TfPavNMT and CjCNMT, respectively. Although buried surface area is similar in GfTNMT, the occurrence of several salt bridges renders the dimerization substantially more favorable, suggesting that the importance of dimerization may vary across BIA NMTs. In all cases, dimerization occurs at the “rear” of the monomer with respect to the substrate-binding pocket and *via* somewhat conserved residues located primarily in two helices and two beta sheets distributed between the N-terminal extension and C-terminal BIA-binding domain. Notably, reciprocal interactions between tri-lysine motifs present in the C-terminus of all BIA NMTs also contribute to the dimerization interface. In Arabidopsis and other eukaryotes, such a motif has been shown to result in retrograde trafficking (ER retention) of membrane-bound proteins *via* interaction with COPI proteins (Wang et al., 2011). However, given that BIA NMTs have been experimentally shown to localize to the cytosol, it appears that this motif may have a more general utility in enabling protein–protein interaction (Hagel and Facchini, 2012). Neither the *in planta* occurrence or functional significance of homo- or hetero-dimerization has been verified for BIA NMTs. However, it has been speculated that small hinge movements in one monomer might be transferred *via* the dimerization domain to the other monomer, resulting in a cooperativity effect. Nevertheless, in the absence of close interactions between the dimerization domain and catalytic site, it would be surprising to discover consequences as substantial as those discussed for the OMTs.

SAM Binding

The Rossmann fold SAM-binding domains of various BIA NMTs are structurally quite similar, yielding RMSD values of ~0.4–0.5 Å when aligned to each other (Torres et al., 2016; Lang et al., 2019). Binding of the cosubstrate occurs *via* sequence motifs largely conserved with other plant SAM-dependent MTs (Kozbial and Mushegian, 2005; Gana et al., 2013). In the available crystal structures, up to 13 direct or water-mediated hydrogen bonds appear to position the cosubstrate (Figure 4). The implicated residues are almost perfectly conserved in all cloned BIA NMTs, with the exception of one non-conservative substitution (Gln to His) in PsRNMT (Supplementary Figure 6) (Torres et al., 2016; Bennett et al., 2018; Lang et al., 2019). Comparison of the apoenzyme and binary complexes (e.g., TfPavNMT *versus* TfPavNMT + SAH) revealed only a minor hinge movement of domains relative to that seen in the OMTs. Instead, active site closure seems to depend on the gate-like 70s loop, which becomes more ordered upon cosubstrate binding. Sequence identity in these features strongly suggests that the mechanisms of SAM binding are conserved for all known BIA NMTs.

BIA Binding

As expected, given the diversity of substrates turned over by various BIA NMTs, the largely helical BIA binding domain shows substantially less conservation than the SAM-binding domain. Although crystal forms binding the enzymes' preferred or physiological substrates (e.g., reticuline or pavine for TfPavNMT, coclaurine for CjCNMT, stylopine for GfTNMT) have been elusive, structures with bound analogs (e.g., *N*-methylheliamine, tetrahydropapaverine), nevertheless, provide some insight into substrate recognition and catalysis. Fortunately, a GfTNMT crystal was obtained in complex with the endogenous product (*S*)-*cis*-*N*-methylstylopine (SMS), providing the most reliable view to date of biologically meaningful interactions. In all BIA NMTs, the binding pocket is lined with many hydrophobic residues and apparently gated by a loop, which becomes ordered and partially helical upon substrate binding. While a limited number of hydrogen bonds are recognizable in the GfTNMT crystal structure, as a general rule, it appears that substrate recognition and binding depends primarily on steric effects, Van der Waals interactions, and aromatic interactions (Figure 4) (Torres et al., 2016; Lang et al., 2019).

The availability of structures for three BIA NMTs reveals that the key substrate-binding interactions differ somewhat across functional subtypes. In TfPavNMT, 13 residues line the binding pocket, and the most significant substrate interactions involve Phe96, His232, Phe292, and Phe330. In CjCNMT, three residues lining one side of the binding pocket (Tyr328, Trp329, Phe332) were probed by site-directed mutagenesis and shown to significantly impact function (Bennett et al., 2018). Characterization of the mutant enzymes' kinetics with respect to isoquinoline and benzylisoquinoline substrates revealed that Trp329 primarily interacts with the benzyl moiety, whereas Phe332 primarily interacts with the isoquinoline moiety. In GfTNMT, comparable results were obtained in which substitution of the equivalent residues (Ile329 and Phe332) with alanine resulted in mutant enzymes with roughly 10% activity (Lang

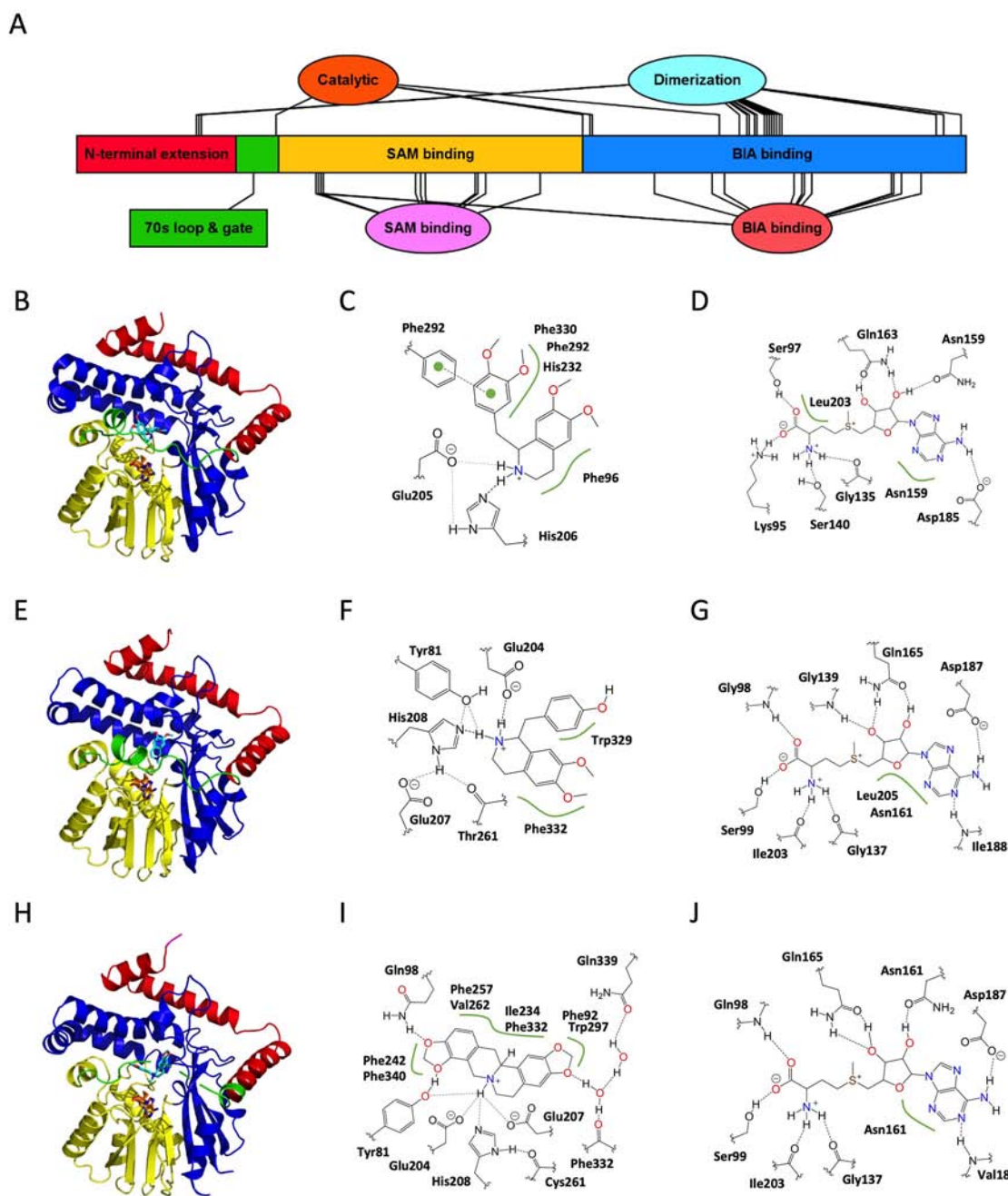


FIGURE 4 | Structure and major active site interactions of representative BIA *N*-methyltransferases. **(A)** Key features of BIA NMTs are labeled according to their relative location along the *T. flavum* pavine *N*-methyltransferase (TfPavNMT) polypeptide with residues denoted by ovals and larger features by rectangles. **(B)** Crystal structures of TfPavNMT with bound SAH and (*R,S*)-tetrahydropapaverine are reported by Torres et al. (2016) (PDB 5KOK). **(E)** *C. japonica* coclaurine NMT with bound SAH and *N*-methylheliamine is reported by Bennett et al. (2018) (PDB 6GKV), and **(H)** *G. flavum* tetrahydropapaverine NMT with bound SAH and (*S*)-*cis*-*N*-methylstylopine is reported by Lang et al. (2019) (PDB 6P3O). Domains correspond to those shown in **(A)** and **Supplementary Figure 6**. The alkaloid substrate is shown in cyan, and the cosubstrate SAH is shown in orange. Major active site interactions in **(C, D)** TfPavNMT, **(F, G)** CjCNMT, and **(I, J)** GfTNMT were visualized using PoseView and modified where necessary to reflect literature reports. Dashed lines indicate hydrogen bonds, and green lines indicate hydrophobic or aromatic interactions. Dashed lines drawn in gray indicate multiple alternative potential interactions. **(A)** Drawn using Illustrator for Biological Sequences (Liu et al., 2015).

et al., 2019). While GfTNMT's lack of activity with isoquinoline substrates precluded a comparative kinetic analysis such as was carried out for CjCNMT, examination of the GfTNMT crystal

structure allowed for the authors to suggest that Phe332 is better positioned to interact with the benzyl moiety and associated C9/C10 methylenedioxy bridge of SMS.

In contrast to the other structurally characterized BIA NMTs, GfTNMT does appear to form a limited set of hydrogen bonds with the BIA molecule's functional groups. Whereas the side chains of residues Gln98 and Tyr81 interact directly with the oxygen atoms of the C9/C10 methylenedioxy bridge in SMS, residues Phe332 (main chain carbonyl) and Gln339 (side chain carboxamide) interact with the C2/C3 oxygen atoms *via* a network of water-mediated hydrogen bonds. Substitution of Tyr81 with either phenylalanine or alanine resulted in comparable mutant enzymes displaying 10–20% activity, indicating that the hydroxyl group of the tyrosine side chain is crucial to the residue's function (Lang et al., 2019). While the involvement of water molecules in substrate binding was only explicitly reported for GfTNMT, careful examination of the TfPavNMT and CjCNMT crystal structures reveals the existence of several well-ordered water molecules within potential hydrogen bonding distance of the BIA molecule.

Interestingly, the size of the substrate-binding pocket in TfPavNMT is somewhat larger than that reported for the other BIA NMTs (Torres et al., 2016). Enlargement of the cavity apparently results from many small contributions, including individual amino acid substitutions (e.g., A204), rotation of side chains and slight repositioning of secondary structural elements (e.g., helix α_4 , 240s loop). Although two separate molecules of a 1-BIA substrate were bound in the reported structure, it was suggested that BIA dimers (bisbenzylisoquinolines) common in *Thalictrum* spp. might bind to the active site *in planta*. While bis-BIAs might conceivably be substrates for *N*-methylation, their status as pathway end products makes it tempting to speculate that they could instead act as feedback inhibitors in a manner analogous to that described for sanguinarine and Tf6OMT. Alternatively, binding of a second BIA monomer at the adjacent (non-productive) location in the active site might cause either substrate or product inhibition and thus regulate NMT activity and adjust pathway flux. Given that binding of a second 1-BIA molecule was proposed to be inhibitory in TfPavNMT, it is reasonable to suspect that the active site might be structured differently when a single substrate molecule is bound. One likely possibility is a tighter interaction with the 70s loop “gate”, which is notably displaced in the available TfPavNMT structure relative to CjCNMT.

In sum, the substrate-binding interactions presently described in the literature vary quite a bit despite substantial conservation of active site residues and domain structure. Given that crystal structures are only available for one representative of each BIA NMT subtype (e.g., TNMT, CNMT, RNMT/PavNMT), it is not presently clear whether the details of substrate binding are conserved across functionally analogous enzymes from different species (e.g., TfCNMT, CjCNMT, PsCNMT, GfNMT1). However, comparably low levels of sequence conservation (i.e., 45–80% identity) make it reasonable to suspect that the specific interactions might be variable. Accordingly, attempts to generalize from available structural information may be misleading, and additional studies are warranted.

70s Loop “Gate”

The 70s loop or active site “gate” mentioned above is a particularly intriguing feature that is not well understood. This region of the

polypeptide undergoes a transition from disordered to ordered form (including structuration of helix α_4) upon SAM and BIA binding, yet no direct interactions are made with the substrate (Torres et al., 2016; Bennett et al., 2018). Nevertheless, patterns of sequence conservation within BIA NMT subtypes (e.g., CNMTs *versus* TNMTs) strongly suggest that the gate contributes to functional differences (**Supplementary Figure 6**). To date, no mutagenesis studies have examined the functional consequences of these variable residues.

On the other hand, two residues adjacent to the 70s loop, which show perfect conservation among the BIA NMTs have been mutagenized with dramatic outcomes. Unexpectedly, replacement of Glu80 in TfPavNMT with alanine resulted in a substantial increase in activity, which was especially notable for non-endogenous substrate (*R,S*)-tetrahydropapaverine (Torres et al., 2016). In GfTNMT, mutagenesis of the equivalent Glu82 leads to a decrease in all activities. However, the effect was comparable in the sense that the mutant enzyme's substrate preference was shifted in favor of scoulerine, which is not considered to be an endogenous substrate for TNMT *in planta*. Examination of crystal structures shows that Glu80/Glu82 hydrogen bonds with an adjacent helix and suggests that the residue might contribute to substrate selectivity by anchoring the 70s loop. The second adjacent residue, Tyr79/Tyr81, is less consistently positioned across BIA NMTs, and mutagenesis also had variable effects depending on the enzyme–substrate pair. In CjCNMT and GfTNMT, the tyrosyl side chain points into the active site and might directly interact with the BIA amino group or benzyl moiety functional groups (**Figure 4**). GfTNMT mutants in which Tyr81 was replaced with alanine, phenylalanine, or arginine all showed substantially reduced activities, indicating that a rather specific interaction (likely involving a hydrogen bond) takes place in the wildtype enzyme. In TfPavNMT, the Tyr79 side chain is rotated approximately 45° with reference to the other BIA NMT structures and does not directly interact with the substrate. Interestingly, replacement of this residue with alanine almost entirely abolished activity with a potential endogenous substrate ((*S*)-reticuline) but activity with the non-endogenous (*R,S*)-tetrahydropapaverine was greatly enhanced. Together, these pieces of evidence suggest that Tyr79/Tyr81 also contributes to selectivity for particular substrates. Thus, it is clear that the 70s loop “gate” and adjacent residues have important consequences regarding BIA NMT activity. While it appears that the outcome of these contributions is substrate selectivity, the precise mechanisms by which this occurs remain to be elucidated.

Catalysis

The mechanisms of catalysis in BIA NMTs are not yet entirely resolved, but appear similar in some ways to that described above for BIA OMTs (Torres et al., 2016; Bennett et al., 2018; Lang et al., 2019). Perfectly conserved glutamic acid and histidine residues are implicated in the methyl transfer reaction, although their contributions may vary from one enzyme–substrate pair to the next. The histidine residue is proposed to act as a general base, which deprotonates the substrate nitrogen and thus activates it to carry out nucleophilic attack on the labile methyl group of SAM (Bennett et al., 2018). In support of an important role for

the histidine residue, substitution of His206 in TfPavNMT and His208 in CjCNMT with alanine resulted in mutant enzymes with sharply reduced activities. Nevertheless, activity was not entirely abolished, and so, it appears that, unlike in BIA OMTs, deprotonation by a general base mechanism is not strictly necessary for catalysis.

Several proposals have been put forth concerning the function of the adjacent glutamic acid residue. In analogy to the catalytic dyad of BIA OMTs (i.e., histidine and aspartic acid), the glutamic acid carboxyl group might hydrogen bond with the histidine imidazole moiety and thus promote the basicity required for substrate deprotonation. However, given that mutagenesis of Glu207 in CjCNMT had only a minor effect on activity (~35% decrease), it was proposed that the backbone carbonyl of another highly conserved residue (Thr261 in CjCNMT) might be the catalytic partner. Another possible mechanism for the glutamic acid is hydrogen bonding with the BIA nitrogen atom to improve its position and orientation relative to the incoming methyl group. Similarly, the glutamic acid might hydrogen bond with the methyl group, itself, in a manner that stabilizes the reaction intermediate, as has been proposed for rat glycine NMT on the basis of computational modeling (Świderek et al., 2018). Aside from hydrogen bond-mediated interactions, it has also been proposed that the reaction intermediate could be stabilized by electrostatic interactions with the negatively charged glutamic acid side chain. Experimental results, in which double mutants (e.g., TfPavNMT-Glu205Ala-His206Ala) show an additive effect, are consistent with any of these alternative models in which the key active site residues make independent contributions to catalysis.

Notably, a general base mechanism may not be applicable to catalysis of all substrates accepted by BIA NMTs. For example, the tertiary amine in GfTNMT substrate stylopine has a calculated pKa of approximately 5.3, which indicates that it would be almost entirely deprotonated under physiological pH conditions, and thus, proton abstraction would be unnecessary to allow methyl transfer (www.chemicalize.org). Mutagenesis of His208 and Glu207 in GfTNMT had effects comparable to those seen in other BIA NMTs (for which substrates almost certainly require deprotonation), indicating that these residues are important for catalysis even when a general base mechanism cannot be invoked.

A fourth residue showing intriguing patterns of conservation between BIA NMT subtypes has also been shown to interact with the substrate (Glu204 in CjCNMT) (Bennett et al., 2018). Due to the side chain's proximity to the alkaloid nitrogen atom and the substantial detrimental effect of alanine substitution on catalysis, Micklefield and colleagues proposed that Glu204 hydrogen bonds with, and helps fine-tune the position and reactivity of, the target nitrogen atom. Interestingly, alanine substitution of Glu204 in GfTNMT had a comparable effect on activity with stylopine despite this molecule's nitrogen atom being unable to form an equivalent hydrogen bond concurrent with methyl transfer. Accordingly, in GfTNMT, it appears more likely that Glu204 interacts with the transition state methyl group or contributes an electrostatic effect. Given that the equivalent residue in TfPavNMT is an alanine, it is clear that the details of this interaction are not conserved between BIA NMT functional subtypes. Intriguingly, RNMT-like enzymes, which efficiently

catalyze *N*-methylation of tertiary BIAs (e.g., reticuline) have a glycine at this position (Morris and Facchini, 2016). The absence of a side chain likely alleviates steric hindrance, which would be expected to occur between the *N*-methyl group and the glutamic acid present in CNMTs. Mutational studies, as well as investigation of natural variation of this position in diverse BIA NMTs, should allow for this hypothesis to be tested.

Although the above proposals likely explain a significant portion of BIA NMT catalytic power, experimental results clearly show that no residue identified to date (Glu204, Glu207, His208, and equivalents) is strictly necessary for methyl transfer to occur. This is in notable contrast to mutagenesis studies on BIA OMTs in which substitution of the general base histidine precludes all activity (Park et al., 2018; Cabry et al., 2019). In fact, even the simultaneous replacement of up to three putative catalytic residues with alanine produced mutant TfPavNMT and GfTNMT enzymes with detectable activities (Torres et al., 2016; Lang et al., 2019). While it remains possible that a general base mechanism is at play in certain reactions, the aforementioned work strongly suggests that rate enhancement in BIA NMTs results from multiple modest contributions. Mechanisms consistent with this interpretation include electrostatic or hydrogen bond-mediated stabilization of the reaction intermediate, and those which invoke "compression, proximity, orientation and desolvation" effects (Zubieta et al., 2003). While the identification and mutagenesis of additional catalytic residues may assist in resolving this question, biophysical modeling may be more likely to provide a conclusive answer.

EVOLUTION OF DIVERSITY IN BIA METHYLTRANSFERASES

Compared to our detailed understanding of the evolutionary trajectories of MT activities implicated in caffeine biosynthesis, relatively little equivalent knowledge is presently available concerning the BIA MTs. Nevertheless, the recent rapid increase in availability of sequence, structure, and function information for these important enzymes foreshadows a commensurate leap forward in our understanding. Given the substantially more complex pathways involved in BIA biosynthesis, understanding the evolution of even a subset of the enzymes will greatly enhance our conceptual grasp of how the tremendous chemodiversity in plants originates.

O-Methyltransferases

The importance of *O*-methylation in various aspects of plant metabolism (e.g., lignin, phenylpropanoid, phytoalexin, and phytohormone biosynthesis) makes it unsurprising that their classification and evolutionary relatedness has received substantial attention. Setting aside the highly sequence-divergent SABATH MTs, the remainder of plant OMTs are generally understood to fall within two major groups (Lam et al., 2007). The first (class I) contains mostly enzymes, which target hydroxycinnamoyl-CoA esters and participate in lignin biosynthesis as well as carboxylic acid OMTs involved in plant hormone and scent metabolism, and the second (class II) contains a much more variable set of enzymes

including those involved in phenylpropanoid and alkaloid biosynthesis. The first group is thought to have evolved as a result of pressures relating to the colonization of terrestrial habitats, whereas the second group likely arose later in response to a more diverse set of evolutionary forces. Although a focused and up-to-date phylogenetic study of plant OMTs is overdue, basic analyses reported in conjunction with the isolation of new alkaloid OMTs have consistently placed the BIA OMTs within class II (Morishige et al., 2002; Morishige et al., 2010; Nomura and Kutchan, 2010; Dang and Facchini, 2012; Salim et al., 2018). While a complete examination of known plant OMTs is outside the scope of this review, we performed a limited analysis of representative plant OMTs, BIA OMTs, and other recently reported alkaloid OMTs to place them in the context of the phylogeny reported by Dananyandan and colleagues (Figure 5) (Lam et al., 2007). Consistent with the previous work, our results place all cloned BIA OMTs within a well-supported clade corresponding to class II. We recovered four subclades with topology and bootstrap support similar to those reported previously. A large majority of BIA OMTs fell within subclade II-D, which was sister to subclade II-C in which most other alkaloid OMTs and flavonoid OMTs clustered. The functionally validated BIA OMTs in subclade II-D appear to be monophyletic, and most prefer substrates of the simple 1-BIA structure (Supplementary Table 1). In addition, the heterodimer-forming OMTs participating in noscapine biosynthesis are within this clade, with the two functionally interchangeable monomers (PsOMT3 and Ps6OMT) clustering together. A single putative BIA OMT, for which preliminary reports indicate a unique ability to methylate the free hydroxyl group of 10-hydroxydihydrosanguinarine, falls within clade II-C in a position distant from the other BIA OMTs and apparently more closely related to other alkaloid OMTs (Purwanto, 2017). Many of the remaining BIA OMTs belong to the very well-supported subclade II-A, which also included hydroxycinnamic acid OMTs involved in phenylpropanoid biosynthesis. All of the BIA OMTs in this subclade are functionally characterized as preferring or accepting protoberberine substrates. The smallest and least well-supported subclade (II-B) contained a few remaining BIA OMTs that mostly target the 7-hydroxyl of simple 1-BIAs, along with OMTs involved in flavonoid and methoxypyrazine biosynthesis.

As suggested in previous publications, the presence of several well-separated clades makes it reasonable to suspect that BIA OMTs have evolved repeatedly in plants. However, it should be noted that the various clades of BIA OMTs show distinct substrate- and regio-specificity and thus cannot be said to have converged on precisely the same function (Figure 5). Although the most commonly accepted explanation at present is repeated evolution, it remains possible that an ancestral OMT was broadly promiscuous and had activity with BIAs, which was lost in most non-BIA OMT lineages. While this may be less likely given the multitude of non-BIA plant OMTs in existence, certain pieces of evidence suggest that it is a question worth exploring. Phenylpropanoid OMTs cloned from *T. tuberosum* were shown to also accept benzylisoquinoline substrates and, similarly, enzymes implicated in the biosynthesis of ipecac alkaloids in *Psychotria ipecacuanha* catalyze methylation of

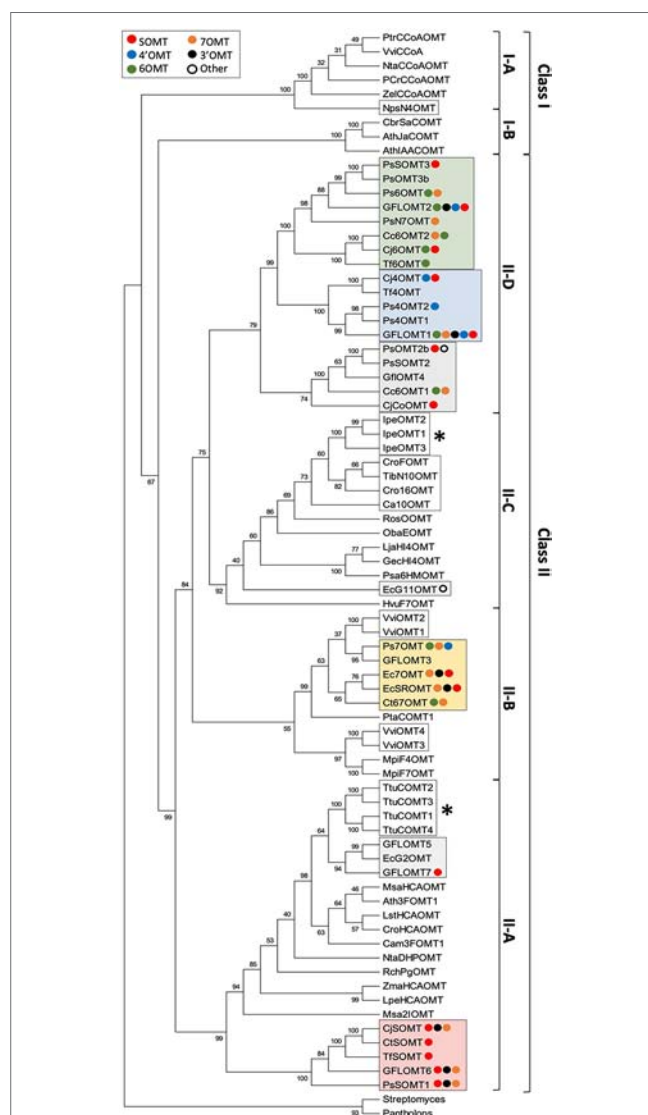


FIGURE 5 | Rooted neighbor-joining phylogenetic tree of BIA O-methyltransferases and representative plant O-methyltransferases labeled with reported *in vitro* substrate range. Frequencies shown at each node represent the percentage of 500 bootstrapped replicate trees in which the associated taxa clustered together. The consensus tree is shown without branch lengths. The analysis involved 77 sequences, and positions with less than 50% coverage were discarded, resulting in a final dataset with 353 positions. The tree was rooted with sequences from bacteria and animals. The analysis was conducted in MEGA7 (Kumar et al., 2016). Representative plant OMT sequences and clade assignments were adapted from Lam et al. (2007). Plant OMTs implicated in the biosynthesis of alkaloids other than BIAs are indicated by white boxes, whereas those that nevertheless accept BIA substrates are indicated with asterisks. BIA OMTs are labeled with colored circles representing their reported *in vitro* activities, in order from strongest to weakest (red, scoulerine-O-methylation; blue, 4'-O-methylation; green, 6-O-methylation; yellow, 7-O-methylation; black, 3'-O-methylation; white, other). Clades with recognizable majority activities are shaded with rectangles in the corresponding color. Genbank accession numbers for BIA OMTs are provided in **Supplementary Table 1**. Other alkaloid OMT accession numbers are: NpsN4'OMT (KJ584561), lpeOMT1 (AB527082), lpeOMT2 (AB527083), lpeOMT3 (AB527084), Ca1OMT (MG996006), Cro16OMT (EF444544), TibN10OMT (MH454075), VviOMT1 (KC533529), VviOMT2 (KC533535), VviOMT3 (KC517470), and VviOMT4 (KC517475).

BIAs with surprising efficiency (Frick and Kutchan, 1999; Nomura and Kutchan, 2010). In fact, the highly divergent and distantly related rat liver catechol OMT is also known to accept BIA substrates (Meyerson et al., 1979). Conversely, none of the BIA OMTs assayed with phenylpropanoid substrates have been shown to accept them (Morishige et al., 2002; Ounaroon et al., 2003; Pienkny et al., 2009), perhaps indicating that such a function is the more derived character. It is intriguing to consider that our present understanding of most plant OMTs as relatively specialized for one class of substrate or another might be a result in part from the practical challenges of assaying for many diverse activities when a new enzyme is discovered. Resolving this important question will require that, going forward, workers begin to routinely examine plant OMTs with respect to a wider range of potential and even physiologically unlikely substrates.

Superimposition of reported *in vitro* activities over apparent phylogeny reveals conservation of function in some BIA NMT clades but not others (Figure 5). The existence of one well-separated subclade containing only enzymes, which preferentially methylate protoberberines such as (S)-scoulerine suggests that most SOMTs share a monophyletic origin. In all but one case, members of this subclade have also been shown to methylate simple 1-BIA substrates to a lesser degree. While this might be interpreted as maintenance of an ancestral enzyme feature, structural comparison of the flexible 1-BIAs to the rather rigid protoberberines suggests that the former can readily adopt a conformation mimicking the latter, thus potentially explaining the functional overlap from a strictly structural point of view. It is interesting to note that the hydroxyl groups methylated by SOMT-like enzymes in simple 1-BIAs (i.e., 7, 3') correspond to those methylated in protoberberines (i.e., 2, 9) (Supplementary Figure 4). Enzymes with SOMT-like activity are also present in other clades. In particular, the *C. japonica* enzyme shown to prefer columbamine over scoulerine may have evolved independently. Clades with majority preferences for 6-, 7-, or 4'-O-methylation of simple 1-BIAs show substantially more functional diversity between members. Although clades showing a preference for 6- or 7-O-methylation are recognizable, most enzymes catalyze both reactions. As mentioned above with reference to the Tf6OMT structure, only a minor adjustment of binding angle is necessary to position either the 6- or 7-hydroxyl in a productive alignment for methyl transfer (Figure 3). Despite their present functional overlap, the distinct clades suggest that 6/7-O-methylation evolved in two independent lineages. Like the SOMTs, enzymes catalyzing 4'-O-methylation primarily fall within a single clade indicative of monophyletic origin. Weak 4' OMT activities also reported for one 6OMT and one 7OMT enzyme may reflect the ability of a simple 1-BIA to bind in a "flipped" orientation as hypothesized in the structural section above. Notably, a clade corresponding primarily to 3'-O-methylation is not evident and such enzymes are present in most clades. Although the above evidence suggests that specialization for various BIA OMT activities occurred in several independent lineages, the sporadic occurrence of corresponding but weaker activities in other clades suggests the possibility that an ancestral BIA

OMT was highly promiscuous and perhaps able to catalyze the full range of methylations with lower efficiency. Resurrection and functional characterization of ancestral BIA OMTs should help test this hypothesis in the near future.

N-Methyltransferases

N-methyltransferases in plants have received less attention overall, and evolutionary relationships are still unclear. It is presently thought that they are polyphyletic in origin and, in fact, that several distinct NMTs may have existed in the last universal common ancestor of all extant life (Anantharaman et al., 2002). Other than the BIA NMTs, major small molecule NMT families in plants include the putrescine NMTs, phosphoethanolamine NMTs, xanthine (SABATH) NMTs, and tocopherol C-methyltransferase-like NMTs (Hibi, 1994; Kato et al., 2000; Nuccio et al., 2000; Liscombe et al., 2010). Given that a comprehensive classification and phylogenetic analysis of plant NMTs is far outside the scope of this review, only the BIA NMT-like enzymes will be considered below.

Phylogenetic analyses carried out in conjunction with the isolation of new BIA NMTs have generally agreed upon the existence of several clades roughly corresponding to subtypes (i.e., CNMT-, RNMT-, or TNMT-like; Figure 6) (Liscombe and Facchini, 2007; Liscombe et al., 2009; Morris and Facchini, 2016). However, the detailed topology of BIA NMT gene trees has varied substantially from one report to the next and all should be interpreted with caution. Recently, an analysis of

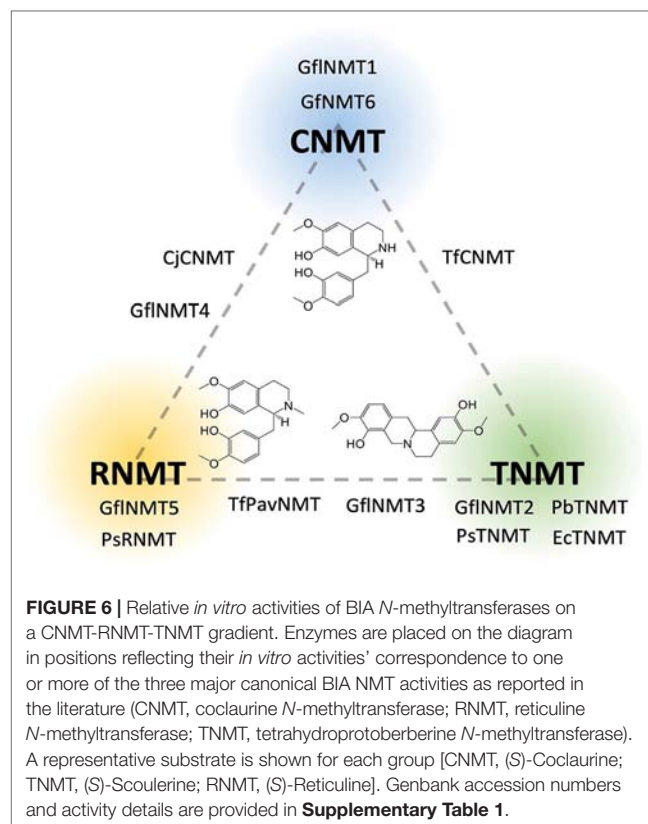


FIGURE 6 | Relative *in vitro* activities of BIA N-methyltransferases on a CNMT-RNMT-TNMT gradient. Enzymes are placed on the diagram in positions reflecting their *in vitro* activities' correspondence to one or more of the three major canonical BIA NMT activities as reported in the literature (CNMT, coclaurine N-methyltransferase; RNMT, reticuline N-methyltransferase; TNMT, tetrahydroprotoberberine N-methyltransferase). A representative substrate is shown for each group [CNMT, (S)-Coclaurine; TNMT, (S)-Scoulerine; RNMT, (S)-Reticuline]. Genbank accession numbers and activity details are provided in **Supplementary Table 1**.

more than 90 putative BIA NMTs from Ranunculales recovered four well-supported clades, three of which correspond to the known BIA NMTs and were experimentally validated as loosely predicting function (Hagel et al., 2015). Superimposition of reported *in vitro* activities over a BIA NMT phylogeny clarifies this idea (Figure 7). In particular, a well-supported monophyletic group of enzymes shown to almost exclusively catalyze *N*-methylation of protoberberine substrates is evident. However, enzymes with weak TNMT-like activities are present in other clades and may have evolved this function independently. Enzymes primarily accepting BIAs with tertiary nitrogen atoms other than protoberberines (i.e., RNMT-like)

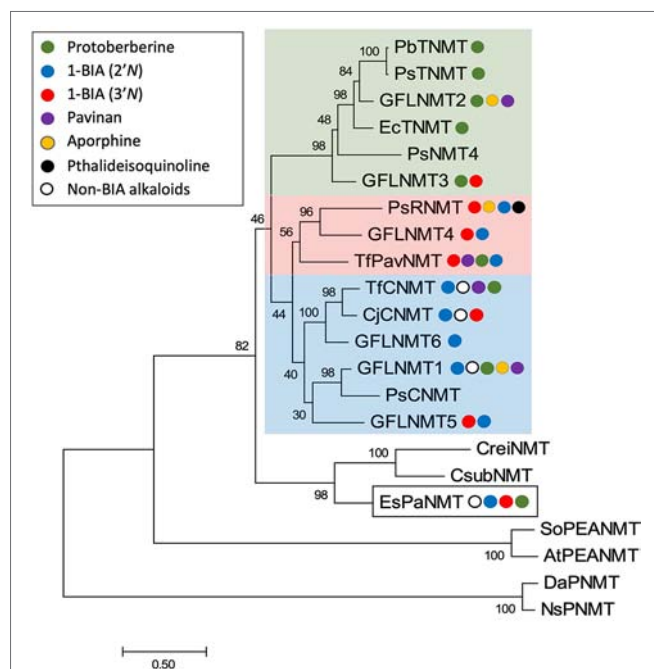


FIGURE 7 | Rooted neighbor-joining phylogenetic tree of BIA *N*-methyltransferases labeled with reported *in vitro* substrate range. The optimal tree is drawn to scale with branch length in units of substitutions per site. Frequencies shown at each node represent the percentage of 500 bootstrapped replicate trees in which the associated taxa clustered together. The analysis involved 22 sequences and positions with less than 50% coverage were discarded, resulting in a final dataset with 358 positions. The tree was rooted with distantly related plant putrescine and phosphoethanolamine *N*-methyltransferase sequences. The analysis was conducted in MEGA7 (Kumar et al., 2016). BIA NMTs are labeled with colored circles representing their reported *in vitro* activities, in order from strongest to weakest (green, protoberberine *N*-methylation; blue, 2' 1-BIA *N*-methylation; red, 3' 1-BIA *N*-methylation; purple, pavinan *N*-methylation; yellow, aporphine *N*-methylation; black, pthalideisoquinoline *N*-methylation; white, *N*-methylation of other alkaloids including isoquinolines). Clades with recognizable majority activities are shaded with rectangles in the corresponding color. The BIA NMT-like *E. sinica* phenylalkylamine NMT (EsPaNMT) implicated in ephedrine biosynthesis is indicated with a white box. Genbank accession numbers for BIA NMTs are provided in **Supplementary Table 1**. *Datura stramonium* putrescine NMT (DaPNMT; CAE47481); *Nicotiana sylvestris* putrescine NMT (NsPNMT; BAA74544); *Arabidopsis thaliana* phosphoethanolamine NMT (AtPEANMT; NP_188427); *Spinacea oleracea* phosphoethanolamine NMT (SoPEANMT; Q9M571). *Coccomyxa subellipsoidea* NMT (CsubNMT; XP_005645141); *Chlamydomonas reinhardtii* NMT (CreiNMT; XP_001695187).

also form a single clade. Notably, members of this clade accept BIAs with a broad range of carbon skeletons, which includes aporphines, pavinans, and pthalideisoquinolines. On the other hand, the cluster of CNMT-like enzymes that preferentially target 1-BIA substrates with secondary nitrogen atoms are reported to have a more restricted substrate range. Although not evident in the phylogenetic analysis presented here, other reports have consistently indicated that CNMT enzymes are more ancestral. Later evolution of the RNMTs and TNMTs is consistent with the cumulative hypothesis, in which enzymes operating further downstream in biosynthetic pathways are recruited later (Granick, 1957). Conclusive statements regarding the evolutionary history of this enzyme family await careful phylogenetic study, ideally including sequences obtained from many species beyond those typically used as model systems for BIA biosynthesis and supported by functional characterization of resurrected ancestral enzymes.

Intriguingly, BLAST searches of publicly available nucleotide sequence databases (NCBI NR, OneKP) reveal that transcripts encoding BIA NMT-like proteins (40–70% amino acid identity; **Supplementary Data Sheet1**) are present in a wide range of flowering plants as well as algae, mosses, gymnosperms, and gnetophytes (Matasci et al., 2014). To the best of our knowledge, all but one of the cloned and functionally characterized members of this large NMT family belong to the Ranunculales order and are implicated in BIA biosynthesis. Given their apparently ancient origin and widespread occurrence, including in species not known to produce alkaloids of any sort, the functional significance and maintenance of BIA NMT-like genes through many millions of years of plant evolution is a fascinating mystery.

Several pieces of evidence point to the possibility of a relatively ancient origin for BIA biosynthesis, which likely included the activity of NMTs. The sporadic but widespread distribution of BIA biosynthesis in eudicots, along with detection of the “gateway” norcoclaurine synthase (NCS) activity in a broad range of plants, supports a proposal that the evolution of BIA biosynthesis may have a monophyletic history in angiosperms (Liscombe et al., 2005). In fact, the occurrence of several BIAs in *Gnetum* species (e.g., 8-benzylberberine), and of NCS activity in *Ephedra distachya*, further suggests that the evolutionary origin of BIA biosynthesis may have been at least as ancient as the divergence of the Gnetophytes (Xu and Lin, 1999; Rochfort et al., 2005; Martin et al., 2011). Although no BIA biosynthetic studies have been completed for these plants, in particular, the production of similar BIAs in other species is known to require a CNMT (Hagel and Facchini, 2013). Interestingly, a recent investigation of *Ephedra sinica* identified a BIA NMT-like enzyme (Phenylalkylamine NMT; EsPaNMT) implicated in the biosynthesis of ephedrine, which also has promiscuous activity on several other alkaloids including 1-BIAs (Morris et al., 2018). This observation supports the notion that the ancestor to all extant BIA NMT-like enzymes may have had a very broad range of activities and which was refined and subfunctionalized in certain lineages, where BIA biosynthesis provided a selective advantage. Of course, an alternative hypothesis is that EsPaNMT and the BIA NMTs were recruited independently from a functionally distinct ancestral lineage and simply converged on BIA NMT function. In any case,

maintenance of BIA NMT-like genes in plants over evolutionary timescales implies that they must function in some useful role. It will be interesting to discover in the coming years whether the annotation of this family as BIA NMTs is simply a historical accident or an accurate representation of their broader roles in plants.

FORCES AND MECHANISMS SHAPING BIA METHYLTRANSFERASE EVOLUTION

As for most specialized metabolites, the forces driving the evolution and maintenance of BIA biosynthesis are not yet fully understood (Weng, 2014). Generally, BIAs are assumed to provide defensive advantages *via* antiherbivore and antimicrobial properties (Hagel and Facchini, 2013). In the case of cultivated BIA-producing varieties, such as *P. somniferum*, artificial selection for the presence of psychoactive morphinans may also have played a minor role in recent times. However, potent biological activities are only firmly demonstrated for a small fraction of BIAs (e.g., berberine, sanguinarine, magnoflorine, morphine), and thus, straightforward adaptive evolution does not comfortably explain the tremendous chemical diversity observed across BIA-producing species and individuals. When attempting to justify the presence of apparently useless, yet metabolically costly, BIAs in a given plant, it is important to appreciate that the biochemical snapshot we obtain in the present day results from a complex evolutionary history spanning innumerable shifts in herbivore and pathogen challenges. Accordingly, biosynthesis of some functionless BIAs may have resulted from pressures no longer present in the environment. Alternatively, it has also been proposed that BIA metabolic diversity may be a useful trait in its own right (Facchini et al., 2004). That is, the production of a large and dynamic repertoire of potential defense molecules, which varies from individual to individual, may represent a form of “diversified bet hedging,” which can ensure the survival of at least some members when a lineage is suddenly faced with novel challenges. The canonical example of this type of coping strategy is seed germination timing, but variation in BIA profile could also conceivably fit the theoretical criteria (i.e., improved long-term evolutionary success despite reduced mean fitness, *via* a reduction in detrimental temporal fitness variance) (Childs et al., 2010). As reviewed in the preceding sections, the BIA MTs show substantial promiscuity and, perhaps more than any other class of enzyme, greatly expand and diversify the pool of BIAs that are produced.

Duplication followed by sub- or neo-functionalization is thought to be a crucial mechanism underlying the diversification of most eukaryotic gene families, and this also applies to the BIA MTs (Taylor and Raes, 2004). In *P. somniferum*, Graham and colleagues identified a genomic region on which many genes required for noscapine biosynthesis were clustered, including three BIA OMT genes corresponding to PsSOMT2, PsSOMT3, and PsSOMT1, various other BIA pathway genes and many transposable elements (Winzer et al., 2012). Based on sequence homology and intron–exon structure, these OMT genes were suggested to have arisen *via* tandem gene duplication, potentially followed by transposon-mediated cluster rearrangement. More recently, a *P. somniferum* whole

genome was reported, which provides further support for a history of MT gene duplication (Guo et al., 2018). The authors described a relatively recent whole genome duplication as well as more ancient segmental duplications likely to have resulted in new BIA MT gene copies. Aside from MTs in the noscapine cluster reported previously, at least seven additional MT genes are present in their assembly (NCBI BioProject PRJNA435796). Notably, two copies of genes encoding Ps6OMT tightly linked to PsCYP80B1 (*N*-methylcoclaurine 3'-hydroxylase) exist on two separate contigs, suggesting the occurrence of either dispersed duplication or tandem duplication followed by genomic rearrangement. In addition, two copies of genes encoding PsN7OMT are identifiable. Given that both 6OMT and N7OMT are necessary for the biosynthesis of papaverine (Figure 2), which is particularly abundant in *P. somniferum*, it appears that increasing gene dosage is one important mechanism enhancing the contribution of key MTs to BIA biosynthesis. In contrast to the OMTs, only single functional copies of genes encoding PsCNMT, PsTNMT, and PsRNMT are evident in the genome. Although not tightly linked, CNMT and TNMT are located in the same region (roughly 10 MBp apart) of one chromosome. Aside from sub- and neo-functionalization, duplicated genes may often become inactive or pseudogenized. For example, three pseudogene copies of TNMT, within ~30 kb of each other, are reportedly linked to the noscapine cluster (Winzer et al., 2012). In addition, examination of the *P. somniferum* genome suggests that such a fate is quite common for duplicated BIA MTs. BLAST searches of the published assembly reveal many putative pseudogenes, corresponding in particular to 4'OMT2, TNMT, and RNMT. These are generally found in tight clusters indicative of tandem duplication. Although outside the scope of this review, it is clear that linkage of MT genes with those encoding upstream and downstream enzymes is an important contributor to the biosynthesis of BIAs. In addition to making co-inheritance of a useful group of alleles more likely, clustering probably facilitates coordinated gene expression *via* chromatin remodeling. Given the substantial amount of clustering evident in the *P. somniferum* genome, it will be interesting to discover whether similar structures exist in other BIA-producing species and, if so, whether clustering is an ancestral feature or yet another example of convergence.

FUTURE DIRECTIONS

In the preceding sections, we reviewed the wealth of information presently available concerning BIA MT structure, function, and relationship with host plant chemodiversity. Although many important insights have been obtained in recent years, much remains to be done if the sea of information is to yield more widely applicable knowledge and conceptual understanding useful to the field of plant biochemistry as a whole.

In spite of the fact that the role of MTs in the central BIA pathway (i.e., leading to core intermediate (S)-reticuline) is firmly established in model species such as *P. somniferum* and other members of Ranunculales, it would be worthwhile to verify that this knowledge

is applicable in more distantly related BIA producing plants such as the Piperales, Cornales, Laurales, Sapindales, and Proteales (Shulgin and Perry, 2002; Liscombe et al., 2005). Similarly, the role of MTs in the many “unusual” BIA branch pathways has not been investigated. This includes pathways biosynthesizing the rhoeadine alkaloids present in *P. rheas* (Rönsch, 1986), benzylprotoberberines (e.g., Latifolian A) occurring in *Gnetum latifolium* (Rochfort et al., 2005), hexahydrobenzophenanthridines (e.g., Corygaline A) occurring in *Corydalis bungeana* (Gao et al., 2018), as well as the dimeric and trimeric BIAs reported in many species (Schiff, 1991). Biosynthesis of aporphine alkaloids in *Nelumbo nucifera* (Proteales) has recently received some attention; however, most analyses assumed that central BIA biosynthesis is the same as in Ranunculales and, furthermore, that *N. nucifera* OMT and NMT homologs catalyze the same reactions as reported in other species (Menéndez-Perdomo and Facchini, 2018). Going forward, it would be valuable to carry out in these species the same types of studies as were used to firmly establish the routes of biosynthesis in Ranunculales (e.g., labeled tracer feeding, detection of intermediates, and activities). Furthermore, it is crucial to experimentally validate the function of putative BIA MTs when they are discovered. Whereas heterologous expression and *in vitro* assays can readily be applied to proteins originating from these other species, *in planta* approaches (e.g., virus-induced gene silencing, CRISPR-mediated knockout) have been used primarily in *P. somniferum*, and substantial method development may be required before these tools can be brought to bear in a wider context. Taken together, the above experiments would shed light on the question of whether BIA biosynthesis truly is monophyletic and as widely conserved as generally assumed or whether different enzymes and pathways have evolved to converge on BIA biosynthesis by different means.

Despite recent publications that have improved the situation, structures of BIA OMTs accepting a full range of BIA scaffolds (e.g., pthalideisoquinolines) or displaying alternate regio-specificity (e.g., 4'OMT) are missing from the literature. Comparison of these with existing structures would suggest how OMTs discriminate between highly similar molecules, and targeted mutagenesis would then allow for experimental validation of these hypotheses. Alternatively, a focused analysis of multiple functionally analogous BIA OMTs from distinct phylogenetic lineages (e.g., Tf6OMT, Ct6/7OMT, Cc6OMT1) would reveal to what extent the mechanisms of substrate binding and catalysis are conserved. Although these types of comparative studies have been limited by the recalcitrance of plant enzymes to crystallization, modern protein engineering methods such as Surface Entropy Reduction can help overcome these challenges (Cooper et al., 2007; Cabry et al., 2019). From a biotechnological standpoint, crystal structures or docking studies with a wider range of inhibitor molecules (e.g., pathway intermediates or end products) would be valuable in pointing the way to engineering feedback-insensitive variants desirable for industrial applications. Although BIA NMTs are relatively well covered in terms of available structures, certain features are still mysterious. Catalysis is still not fully understood, and this issue is compounded by difficulties in obtaining crystal structures with the enzyme's “true” substrate bound. One promising approach to this problem involves the use of a reactive SAM analog

(S-adenosyl-vinthonine) to form a bisubstrate adduct *in situ*, which remains trapped in the active site of the crystallized NMT (Qu et al., 2016). Another open question, potentially explored *via* biophysical modeling, is how these enzymes' binding pockets successfully discriminate between rather similar BIAs despite seemingly forming very few specific interactions. Similarly, the unique N-terminal extension and active site “gate”, which might contribute to substrate selectivity in BIA NMTs, await careful study. In these cases, domain swap or deletion experiments should yield useful information on their function.

Of particular interest is the biochemical and physiological significance of MT dimerization. Structural elucidation of both functional and non-functional heterodimers (e.g., PsSOMT2:PsSOMT3 vs PsSOMT2:PsN7OMT) might reveal the subtly different interactions, which prevent or allow catalysis on certain substrates. However, targeted mutagenesis will undoubtedly be required to verify such hypotheses. Although only one heterodimer is known to be physiologically relevant at present, this is likely to change with more study. Combinatorial expression of BIA MTs in heterologous systems containing reconstituted BIA pathways is a powerful system with which to search for such interactions. However, it will also be crucial to validate that these heterodimers form in plants and make meaningful contributions to biosynthetic capacity. Ideally, this will be done with a combination of *ex vivo* (e.g., pull down, enzyme assay) and *in vivo* (e.g., FRET, gene knockout) methods. Of course, while considering the occurrence of heterodimerization, it will be crucial to also consider higher-order interactions with other proteins and enzymes that might form BIA metabolons.

Recent interest in understanding BIA biosynthesis in a wider range of plants should soon provide a more diverse set of BIA MTs to study. Modern computational resources and algorithms should allow for robust analysis of all these sequences, resulting in reliable phylogenies clarifying their interrelationships. As DNA synthesis costs continue to decrease, resurrection of ancestral enzymes should become routine and will allow us to answer long-standing questions about MT evolution. For example, it will be fascinating to discover what functions the ancestral class II OMT may have had and what trajectories lead to the extant functional diversity. Similarly, it should be possible to test the long-standing hypothesis that extant BIA NMTs diverged from a CNMT-like ancestor, and whether neo-functionalization or, rather, sub-functionalization then came into play. Along with the analysis of transcripts and encoded enzymes, genome structure will undoubtedly contribute to understanding the mechanics of MT evolution. Although the recently published *P. somniferum* genome has begun to shed such light, the significance of certain features (e.g., clustering) would be more evident if the genomes of additional *Papaver* species, more distantly related BIA producers, and closely related non-producers were available. Comparative genomics should reveal the timing of gene duplications and suggest how selection and drift contributed to the present complement of BIA MTs.

Ultimately, a complete understanding of the determinants of BIA MT function and the evolutionary trajectories that led to the formation of specific enzymes will reveal an important part of how the exquisite BIA biosynthetic pathways came to be. In combination with existing knowledge regarding caffeine

biosynthesis and, eventually, with knowledge concerning the many other alkaloid pathways, these discoveries will allow us to reach satisfactory answers to the long-standing questions of how and why plant-specialized metabolism achieves such tremendous diversity.

AUTHOR CONTRIBUTIONS

JM wrote the manuscript and created the figures. PF edited the final draft of the manuscript.

FUNDING

JM is the recipient of a Natural Sciences and Engineering Research Council of Canada Postgraduate Scholarship. This work was

funded by a Natural Sciences and Engineering Research Council of Canada Discovery Grant to PF.

ACKNOWLEDGMENTS

We would like to thank Dr. Samuel Yeaman and Dr. Qiushi Li for their assistance with visualization and BLAST annotation of the published *Papaver somniferum* genome.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.01058/full#supplementary-material>

SUPPLEMENTARY DATASHEET 1 | A Selection of BIA NMT-like proteins sharing 40–70% amino acid sequence identity.

REFERENCES

- Afendi, F. M., Okada, T., Yamazaki, M., Hirai-Morita, A., Nakamura, Y., Nakamura, K., et al. (2012). KNApSACk family databases: integrated metabolite–plant species databases for multifaceted plant research. *Plant Cell Physiol.* 53, e1. doi: 10.1093/pcp/pcr165
- Anantharaman, V., Koonin, E., and Aravind, L. (2002). Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acids Res.* 30, 1427–1464. doi: 10.1093/nar/30.7.1427
- Argout, X., Salse, J., Aury, J. M., Guiltinan, M. J., Droc, G., Gouzy, J., et al. (2011). The genome of *Theobroma cacao*. *Nat. Genet.* 43, 101–108. doi: 10.1038/ng.736
- Ashihara, H., Monteiro, A. M., Gillies, F. M., and Crozier, A. (1996). Biosynthesis of caffeine in leaves of coffee. *Plant Physiol.* 111, 747–753. doi: 10.1104/pp.111.3.747
- Bennett, M. R., Thompson, M. L., Shepherd, S. A., Dunstan, M. S., Herbert, A. J., Smith, D. R. M., et al. (2018). Structure and biocatalytic scope of coclaurine N-methyltransferase. *Angew. Chemie - Int. Ed.* 57, 10600–10604. doi: 10.1002/anie.201805060
- Biaostoff, S., Brandt, W., and Dräger, B. (2009). Putrescine N-methyltransferase – the start for alkaloids. *Phytochemistry* 70, 1708–1718. doi: 10.1016/j.phytochem.2009.06.012
- Cabry, M., Offen, W. A., Saleh, P., Li, Y., Winzer, T., Graham, I. A., et al. (2019). Structure of *Papaver somniferum* O-methyltransferase 1 reveals initiation of noscapine biosynthesis with implications for plant natural product methylation. *ACS Catal.* 9, 3840–3848. doi: 10.1021/acscatal.9b01038
- Chang, L., Hagel, J. M., and Facchini, P. J. (2015). Isolation and characterization of O-methyltransferases involved in the biosynthesis of glaucine in *Glaucium flavum*. *Plant Physiol.* 169, 1127–1140. doi: 10.1104/pp.15.01240
- Childs, D. Z., Metcalf, C. J. E., and Rees, M. (2010). Evolutionary bet-hedging in the real world: empirical evidence and challenges revealed by plants. *Proc. R. Soc. B Biol. Sci.* 277, 3055–3064. doi: 10.1098/rspb.2010.0707
- Choi, K.-B., Morishige, T., Shitan, N., Yazaki, K., and Sato, F. (2002). Molecular cloning and characterization of coclaurine N-methyltransferase from cultured cells of *Coptis japonica*. *J. Biol. Chem.* 277, 830–835. doi: 10.1074/jbc.M106405200
- Cooper, D. R., Boczek, T., Grelewski, K., Pinkowska, M., Sikorska, M., Zawadzki, M., et al. (2007). Protein crystallization by surface entropy reduction: optimization of the SER strategy. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 63, 636–645. doi: 10.1107/S0907444907010931
- Dang, T. T., and Facchini, P. J. (2012). Characterization of three O-methyltransferases involved in noscapine biosynthesis in opium poppy. *Plant Physiol.* 159, 618–631. doi: 10.1104/pp.112.194886
- Denoeud, F., Carretero-Paulet, L., Dereeper, A., Droc, G., Guyot, R., Pietrella, M., et al. (2014). The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science* 345, 1181–1184. doi: 10.1126/science.1255274
- Desagné-Penix, I., and Facchini, P. J. (2012). Systematic silencing of benzylisoquinoline alkaloid biosynthetic genes reveals the major route to papaverine in opium poppy. *Plant J.* 72, 331–344. doi: 10.1111/j.1365-3113.2012.05084.x
- Dunlevy, J. D., Dennis, E. G., Soole, K. L., Perkins, M. V., Davies, C., and Boss, P. K. (2013). A methyltransferase essential for the methoxypyrazine-derived flavour of wine. *Plant J.* 75, 606–617. doi: 10.1111/tj.12224
- Dunlevy, J. D., Soole, K. L., Perkins, M. V., and Boss, P. K. (2010). Two O-methyltransferases involved in the biosynthesis of methoxypyrazines: grape-derived aroma compounds important to wine flavour. *Plant Mol. Biol.* 74, 77–89. doi: 10.1007/s11103-010-9655-y
- Facchini, P. J., Bird, D. A., and St-Pierre, B. (2004). Can *Arabidopsis* make complex alkaloids? *Trends Plant Sci.* 9, 116–122. doi: 10.1016/j.tplants.2004.01.004
- Facchini, P. J., and Park, S. U. (2003). Developmental and inducible accumulation of gene transcripts involved in alkaloid biosynthesis in opium poppy. *Phytochemistry* 64, 177–186. doi: 10.1016/S0031-9422(03)00292-9
- Farrow, S. C., Kamileen, M., Meades, J., Ameyaw, B., Xiao, Y., and O'Connor, S. (2018). Cytochrome P450 and O-methyltransferase catalyse the final steps in the biosynthesis of the anti-addictive alkaloid ibogaine from *Tabernanthe iboga*. *J. Biol. Chem.* 293, 13821–13833. doi: 10.1074/jbc.RA118.004060
- Frick, S., and Kutchan, T. M. (1999). Molecular cloning and functional expression of O-methyltransferases common to isoquinoline alkaloid and phenylpropanoid biosynthesis. *Plant J.* 17, 329–339. doi: 10.1046/j.1365-3113.1999.00379.x
- Fujii, N., Inui, T., Iwasa, K., Morishige, T., and Sato, F. (2007). Knockdown of berberine bridge enzyme by RNAi accumulates (S)-reticuline and activates a silent pathway in cultured California poppy cells. *Transgenic Res.* 16, 363–375. doi: 10.1007/s11248-006-9040-4
- Gana, R., Rao, S., Huang, H., Wu, C., and Vasudevan, S. (2013). Structural and functional studies of S-adenosyl-L-methionine binding proteins: a ligand-centric approach. *BMC Struct. Biol.* 13, 6. doi: 10.1186/1472-6807-13-6
- Gao, C., Gu, X., Wang, X., Cao, H., Lin, B., Liu, Y., et al. (2018). Corygaline A, a hexahydrobenzophenanthridine alkaloid with an unusual carbon skeleton from *Corydalis bungeana* Turcz. *Org. Biomol. Chem.* 16, 8710–8714. doi: 10.1039/C8OB02194E
- Granick, S. (1957). Speculations on the origins and evolution of photosynthesis. *Ann. N.Y. Acad. Sci.* 69, 292–308. doi: 10.1111/j.1749-6632.1957.tb49665.x
- Guillaumie, S., Ilg, A., Réty, S., Brette, M., Trossat-magnin, C., Decroocq, S., et al. (2013). Genetic analysis of the biosynthesis of 2-methoxy-3-isobutylpyrazine, a major grape-derived aroma compound impacting wine quality. *Plant Physiol.* 162, 604–615. doi: 10.1104/pp.113.218313
- Guo, L., Winzer, T., Yang, X., Li, Y., Ning, Z., He, Z., et al. (2018). The opium poppy genome and morphinan production. *Science* 362, 343–347. doi: 10.1126/science.aat4096
- Hagel, J. M., and Facchini, P. J. (2012). Subcellular localization of sanguinarine biosynthetic enzymes in cultured opium poppy cells. *Vitro Cell. Dev. Biol. - Plant* 48, 233–240. doi: 10.1007/s11627-012-9426-3

- Hagel, J. M., and Facchini, P. J. (2013). Benzyloquinoline alkaloid metabolism: a century of discovery and a brave new world. *Plant Cell Physiol.* 54, 647–672. doi: 10.1093/pcp/pct020
- Hagel, J. M., Morris, J. S., Lee, E.-J., Desgagné-Penix, I., Bross, C. D., Chang, L., et al. (2015). Transcriptome analysis of 20 taxonomically related benzyloquinoline alkaloid-producing plants. *BMC Plant Biol.* 15, 227. doi: 10.1186/s12870-015-0596-0
- He, S. M., Liang, Y.-L., Cong, K., Chen, G., Zhao, X., Zhao, Q.-M., et al. (2018). Identification and characterization of genes involved in benzyloquinoline alkaloid biosynthesis in *Coptis* species. *Front. Plant Sci.* 9, 1–13. doi: 10.3389/fpls.2018.00731
- He, S. M., Song, W. L., Cong, K., Wang, X., Dong, Y., Cai, J., et al. (2017). Identification of candidate genes involved in isoquinoline alkaloids biosynthesis in *Dactylicapnos scandens* by transcriptome analysis. *Sci. Rep.* 7, 9119. doi: 10.1038/s41598-017-08672-w
- Hegnauer, R. (1988). Biochemistry, distribution and taxonomic relevance of higher plant alkaloids. *Phytochemistry* 27, 2423–2427. doi: 10.1016/0031-9422(88)87006-7
- Hibi, N. (1994). Gene expression in tobacco low-nicotine mutants. *Plant Cell* 6, 723–735. doi: 10.1105/tpc.6.5.723
- Huang, C. C., Smith, C. V., Glickman, M. S., Jacobs, W. R., and Sacchettini, J. C. (2002). Crystal structures of mycolic acid cyclopropane synthases from *Mycobacterium tuberculosis*. *J. Biol. Chem.* 277, 11559–11569. doi: 10.1074/jbc.M111698200
- Huang, R., O'Donnell, A. J., Barbolini, J. J., and Barkman, T. J. (2016). Convergent evolution of caffeine in plants by co-option of exapted ancestral enzymes. *Proc. Natl. Acad. Sci.* 113, 10613–10618. doi: 10.1073/pnas.1602575113
- Hult, K., and Berglund, P. (2007). Enzyme promiscuity: mechanism and applications. *Trends Biotechnol.* 25, 231–238. doi: 10.1016/j.tibtech.2007.03.002
- Jin, J. Q., Yao, M. Z., Ma, C. L., Ma, J. Q., and Chen, L. (2016). Natural allelic variations of TCS1 play a crucial role in caffeine biosynthesis of tea plant and its related species. *Plant Physiol. Biochem.* 100, 18–26. doi: 10.1016/j.plaphy.2015.12.020
- Jin, Z., and Xu, X.-H. (2013). “Amaryllidaceae alkaloids,” in *Natural Products* (Berlin, Heidelberg: Springer Berlin Heidelberg), 479–522. doi: 10.1007/978-3-642-22144-6_18
- Junker, A., Fischer, J., Sichart, Y., Brandt, W., and Dräger, B. (2013). Evolution of the key alkaloid enzyme putrescine *N*-methyltransferase from spermidine synthase. *Front. Plant Sci.* 4, 260. doi: 10.3389/fpls.2013.00260
- Kato, M., Kanehara, T., Shimizu, H., Suzuki, T., Gillies, F. M., Crozier, A., et al. (1996). Caffeine biosynthesis in young leaves of *Camellia sinensis*: *in vitro* studies on *N*-methyltransferase activity involved in the conversion of xanthosine to caffeine. *Physiol. Plant.* 98, 629–636. doi: 10.1111/j.1399-3054.1996.tb05720.x
- Kato, M., Mizuno, K., Crozier, A., Fujimura, T., and Ashihara, H. (2000). Caffeine synthase gene from tea leaves. *Nature* 406, 956–957. doi: 10.1038/35023072
- Kato, M., Mizuno, K., Fujimura, T., Iwama, M., Irie, M., Crozier, A., et al. (1999). Purification and characterization of caffeine synthase from tea leaves. *Plant Physiol.* 120, 579–586. doi: 10.1104/pp.120.2.579
- Kilgore, M. B., Augustin, M. M., Starks, C. M., O'Neil-Johnson, M., May, G. D., Crow, J. A., et al. (2014). Cloning and characterization of a norbelladine 4'-*O*-methyltransferase involved in the biosynthesis of the alzheimer's drug galanthamine in *Narcissus* sp. aff. *pseudonarcissus*. *PLoS One* 9, e103223. doi: 10.1371/journal.pone.0103223
- Kozbial, P. Z., and Mushegian, A. R. (2005). Natural history of *S*-adenosylmethionine-binding proteins. *BMC Struct. Biol.* 5, 19. doi: 10.1186/1472-6807-5-19
- Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 372, 774–797. doi: 10.1016/j.jmb.2007.05.022
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054
- Lam, K. C., Ibrahim, R. K., Behdad, B., and Dayanandan, S. (2007). Structure, function, and evolution of plant *O*-methyltransferases. *Genome* 50, 1001–1013. doi: 10.1139/G07-077
- Lang, D. E., Morris, J. S., Rowley, M., Torres, M. A., Maksimovich, V. A., Facchini, P. J., et al. (2019). Structure–function studies of tetrahydroprotoberberine *N*-methyltransferase reveal the molecular basis of stereoselective substrate recognition. *J. Biol. Chem.* doi: 10.1074/jbc.RA119.009214
- Larsson, K. A. E., Zetterlund, I., Delp, G., and Jonsson, L. M. V. (2006). *N*-methyltransferase involved in gramine biosynthesis in barley: cloning and characterization. *Phytochemistry* 67, 2002–2008. doi: 10.1016/j.phytochem.2006.06.036
- Lee, S. G., Kim, Y., Alpert, T. D., Nagata, A., and Jez, J. M. (2012). Structure and reaction mechanism of phosphoethanolamine methyltransferase from the malaria parasite *Plasmodium falciparum*. *J. Biol. Chem.* 287, 1426–1434. doi: 10.1074/jbc.M111.315267
- Levac, D., Cázares, P., Yu, F., and De Luca, V. (2016). A picricine *N*-methyltransferase belongs to a new family of γ -tocopherol-like methyltransferases found in medicinal plants that make biologically active monoterpenoid indole alkaloids. *Plant Physiol.* 170, 1935–1944. doi: 10.1104/pp.15.01813
- Levac, D., Murata, J., Kim, W. S., and De Luca, V. (2008). Application of carborundum abrasion for investigating the leaf epidermis: molecular cloning of *Catharanthus roseus* 16-hydroxytabersonine-16-*O*-methyltransferase. *Plant J.* 53, 225–236. doi: 10.1111/j.1365-3113X.2007.03337.x
- Li, Y., and Smolke, C. D. (2016). Engineering biosynthesis of the anticancer alkaloid noscapine in yeast. *Nat. Commun.* 7, 1–14. doi: 10.1038/ncomms12137
- Liu, W., Xie, Y., Ma, J., Luo, X., Nie, P., Zuo, Z., et al. (2015). IBS: an illustrator for the presentation and visualization of biological sequences. *Bioinformatics* 31, 3359–3361. doi: 10.1093/bioinformatics/btv362
- Liscombe, D. K., and Facchini, P. J. (2007). Molecular cloning and characterization of tetrahydroprotoberberine cis-*N*-methyltransferase, an enzyme involved in alkaloid biosynthesis in opium poppy. *J. Biol. Chem.* 282, 14741–14751. doi: 10.1074/jbc.M611908200
- Liscombe, D. K., Usera, A. R., and O'Connor, S. E. (2010). Homolog of tocopherol *C* methyltransferases catalyzes *N* methylation in anticancer alkaloid biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* 107, 18793–18798. doi: 10.1073/pnas.1009003107
- Liscombe, D. K., MacLeod, B. P., Loukanina, N., Nandi, O. I., and Facchini, P. J. (2005). Evidence for the monophyletic evolution of benzyloquinoline alkaloid biosynthesis in angiosperms. *Phytochemistry* 66, 1374–1393. doi: 10.1016/j.phytochem.2005.04.029
- Liscombe, D. K., Ziegler, J., Schmidt, J., Ammer, C., and Facchini, P. J. (2009). Targeted metabolite and transcript profiling for elucidating enzyme function: isolation of novel *N*-methyltransferases from three benzyloquinoline alkaloid-producing species. *Plant J.* 60, 729–743. doi: 10.1111/j.1365-3113X.2009.03980.x
- Martin, F., Grkovic, T., Sykes, M. L., Shelper, T., Avery, V. M., Camp, D., et al. (2011). Alkaloids from the Chinese vine *Gnetum montanum*. *J. Nat. Prod.* 74, 2425–2430. doi: 10.1021/np200700f
- Matasci, N., Hung, L. H., Yan, Z., Carpenter, E. J., Wickett, N. J., Mirarab, S., et al. (2014). Data access for the 1,000 Plants (1KP) project. *Gigascience* 3, 1–10. doi: 10.1186/2047-217X-3-17
- McCarthy, A., and McCarthy, J. G. (2007). The structure of two *N*-methyltransferases from the caffeine biosynthetic pathway. *Plant Physiol.* 144, 879–889. doi: 10.1104/pp.106.094854
- Menéndez-Perdomo, I. M., and Facchini, P. J. (2018). Benzyloquinoline alkaloids biosynthesis in sacred lotus. *Molecules* 23, 2899. doi: 10.3390/molecules23112899
- Meyerson, L. R., Cashaw, J. L., McMurtrey, K. D., and Davis, V. E. (1979). Stereoselective enzymatic *O*-methylation of tetrahydropapaveroline and tetrahydroxyberberine alkaloids. *Biochem. Pharmacol.* 28, 1745–1752. doi: 10.1016/0006-2952(79)90535-5
- Minami, H., Kim, J.-S., Ikezawa, N., Takemura, T., Katayama, T., Kumagai, H., et al. (2008). Microbial production of plant benzyloquinoline alkaloids. *Proc. Natl. Acad. Sci. U. S. A.* 105, 7393–7398. doi: 10.1073/pnas.0802981105
- Mizuno, K., Kato, M., Irino, F., Yoneyama, N., Fujimura, T., and Ashihara, H. (2003a). The first committed step reaction of caffeine biosynthesis: 7-methylxanthosine synthase is closely homologous to caffeine synthases in coffee (*Coffea arabica* L.) 1. *FEBS Lett.* 547, 56–60. doi: 10.1016/S0014-5793(03)00670-7
- Mizuno, K., Okuda, A., Kato, M., Yoneyama, N., Tanaka, H., Ashihara, H., et al. (2003b). Isolation of a new dual-functional caffeine synthase gene encoding an enzyme for the conversion of 7-methylxanthine to caffeine from coffee (*Coffea arabica* L.) 1. *FEBS Lett.* 534, 75–81. doi: 10.1016/S0014-5793(02)03781-X
- Morishige, T., Dubouzet, E., Choi, K. B., Yazaki, K., and Sato, F. (2002). Molecular cloning of columbamine *O*-methyltransferase from cultured *Coptis japonica* cells. *Eur. J. Biochem.* 269, 5659–5667. doi: 10.1046/j.1432-1033.2002.03275.x
- Morishige, T., Tamakoshi, M., Takemura, T., and Sato, F. (2010). Molecular characterization of *O*-methyltransferases involved in isoquinoline alkaloid biosynthesis in *Coptis japonica*. *Proc. Jpn. Acad. Ser. B. Phys. Biol. Sci.* 86, 757–768. doi: 10.2183/pjab.86.757

- Morishige, T., Tsujita, T., Yamada, Y., and Sato, F. (2000). Molecular characterization of the *S*-adenosyl-L-methionine:3'-hydroxy-*N*-methylcoclaurine 4'-*O*-methyltransferase involved in isoquinoline alkaloid biosynthesis in *Coptis japonica*. *J. Biol. Chem.* 275, 23398–23405. doi: 10.1074/jbc.M002439200
- Morris, J. S., and Facchini, P. J. (2016). Isolation and characterization of reticuline *N*-methyltransferase involved in biosynthesis of the aporphine alkaloid magnoflorine in opium poppy. *J. Biol. Chem.* 291, 23416–23427. doi: 10.1074/jbc.M116.750893
- Morris, J. S., Groves, R. A., Hagel, J. M., and Facchini, P. J. (2018). An *N*-methyltransferase from *Ephedra sinica* catalyzing the formation of ephedrine and pseudoephedrine enables microbial phenylalkylamine production. *J. Biol. Chem.* 293, 13364–13376. doi: 10.1074/jbc.RA118.004067
- Mösl Waldhauser, S. S., Gillies, F. M., Crozier, A., and Baumann, T. W. (1997a). Separation of the *N*-7-methyltransferase, the key enzyme in caffeine biosynthesis. *Phytochemistry* 45, 1407–1414. doi: 10.1016/S0031-9422(97)00187-8
- Mösl Waldhauser, S. S., Kretschmar, J. A., and Baumann, T. W. (1997b). *N*-methyltransferase activities in caffeine biosynthesis: biochemical characterization and time course during leaf development of *Coffea arabica*. *Phytochemistry* 44, 853–859. doi: 10.1016/S0031-9422(96)00662-0
- Navarro, G., and Elliott, H. W. (1971). The effects of morphine, morphinone and thebaine on the EEG and behavior of rabbits and cats. *Neuropharmacology* 10, 367–377. doi: 10.1016/0028-3908(71)90065-7
- Nomura, T., and Kutchan, T. M. (2010). Three new *O*-methyltransferases are sufficient for all *O*-methylation reactions of ipecac alkaloid biosynthesis in root culture of *Psychotria ipecacuanha*. *J. Biol. Chem.* 285, 7722–7738. doi: 10.1074/jbc.M109.086157
- Nuccio, M. L., Ziemak, M. J., Henry, S. A., Weretilnyk, E. A., and Hanson, A. D. (2000). cDNA cloning of phosphoethanolamine *N*-methyltransferase from spinach by complementation in *Schizosaccharomyces pombe* and characterization of the recombinant enzyme. *J. Biol. Chem.* 275, 14095–14101. doi: 10.1074/jbc.275.19.14095
- O'Brien, P. J., and Herschlag, D. (1999). Catalytic promiscuity and the evolution of new enzymatic activities. *Chem. Biol.* 6, R91–R105. doi: 10.1016/S1074-5521(99)80033-7
- Ogawa, M., Herai, Y., Koizumi, N., Kusano, T., and Sano, H. (2001). 7-Methylxanthine methyltransferase of coffee plants. *J. Biol. Chem.* 276, 8213–8218. doi: 10.1074/jbc.M009480200
- Onaroon, A., Decker, G., Schmidt, J., Lottspeich, F., and Kutchan, T. M. (2003). (*R,S*)-Reticuline 7-*O*-methyltransferase and (*R,S*)-norcoclaurine 6-*O*-methyltransferase of *Papaver somniferum* - cDNA cloning and characterization of methyl transfer enzymes of alkaloid biosynthesis in opium poppy. *Plant J.* 36, 808–819. doi: 10.1046/j.1365-313X.2003.01928.x
- Park, M. R., Chen, X., Lang, D. E., Ng, K. K. S., and Facchini, P. J. (2018). Heterodimeric *O*-methyltransferases involved in the biosynthesis of noscapine in opium poppy. *Plant J.* 95, 252–267. doi: 10.1111/tpj.13947
- Pienkny, S., Brandt, W., Schmidt, J., Kramell, R., and Ziegler, J. (2009). Functional characterization of a novel benzylisoquinoline *O*-methyltransferase suggests its involvement in papaverine biosynthesis in opium poppy (*Papaver somniferum* L.). *Plant J.* 60, 56–67. doi: 10.1111/j.1365-313X.2009.03937.x
- Purwanto, R. (2017). Isolation and characterization of novel *O*-methyltransferase involved in benzylisoquinoline alkaloids biosynthesis in *Eschscholzia californica*. [Kyoto (Japan)]: Kyoto University. doi: 10.14989/doctor.k20780
- Purwanto, R., Hori, K., Yamada, Y., and Sato, F. (2017). Unraveling additional *O*-methylation steps in benzylisoquinoline alkaloid biosynthesis in California poppy (*Eschscholzia californica*). *Plant Cell Physiol.* 58, 1528–1540. doi: 10.1093/pcp/pcx093
- Qu, W., Catcott, K. C., Zhang, K., Liu, S., Guo, J. J., Ma, J., et al. (2016). Capturing unknown substrates via *in situ* formation of tightly bound bisubstrate adducts: *S*-adenosyl-vinthonine as a functional probe for adomet-dependent methyltransferases. *J. Am. Chem. Soc.* 138, 2877–2880. doi: 10.1021/jacs.5b05950
- Reid, K. S. C., Lindley, P. F., and Thornton, J. M. (1985). Sulphur–aromatic interactions in proteins. *FEBS Lett.* 190, 209–213. doi: 10.1016/0014-5793(85)81285-0
- Robin, A. Y., Giustini, C., Graindorge, M., Matringe, M., and Dumas, R. (2016). Crystal structure of norcoclaurine-6-*O*-methyltransferase a key rate-limiting step in the synthesis of benzylisoquinoline alkaloids. *Plant J.* 87, 641–653. doi: 10.1111/tpj.13225
- Rochfort, S. J., Towerzey, L., Carroll, A., King, G., Michael, A., Pierens, G., et al. (2005). Latifolians A and B, novel JNK3 kinase inhibitors from the Papua New Guinea plant *Gnetum latifolium*. *J. Nat. Prod.* 68, 1080–1082. doi: 10.1021/np049616i
- Rohde, B., Hans, J., Martens, S., Baumert, A., Hunziker, P., and Matern, U. (2007). Anthranilate *N*-methyltransferase, a branch-point enzyme of acridone biosynthesis. *Plant J.* 53, 541–553. doi: 10.1111/j.1365-313X.2007.03360.x
- Rönsch, H. (1986). "Rhoeadine alkaloids," in *The Alkaloids: Chemistry and Physiology*, (Cambridge, Massachusetts: Academic Press), 1–93. doi: 10.1016/S0099-9598(08)60113-3
- Saito, K., and Matsuda, F. (2009). Metabolomics for functional genomics, systems biology, and biotechnology. *Annu. Rev. Plant Biol.* 61, 463–489. doi: 10.1146/annurev.arplant.043008.092035
- Salim, V., Jones, A. D., and DellaPenna, D. (2018). *Camptotheca acuminata* 10-hydroxycamptothecin *O*-methyltransferase: an alkaloid biosynthetic enzyme co-opted from flavonoid metabolism. *Plant J.* 95, 112–125. doi: 10.1111/tpj.13936
- Samanani, N. (2005). Cell type-specific localization of transcripts encoding nine consecutive enzymes involved in protoberberine alkaloid biosynthesis. *Plant Cell Online* 17, 915–926. doi: 10.1105/tpc.104.028654
- Sato, F., Tsujita, T., Katagiri, Y., Yoshida, S., and Yamada, Y. (1994). Purification and characterization of *S*-adenosyl-L-methionine:norcoclaurine 6-*O*-methyltransferase from cultured *Coptis japonica* cells. *Eur. J. Biochem.* 225, 125–131. doi: 10.1111/j.1432-1033.1994.00125.x
- Schiff, P. L. (1991). Bisbenzylisoquinoline alkaloids. *J. Nat. Prod.* 54, 645–749. doi: 10.1021/np50075a001
- Schimpl, F. C., Kiyota, E., Lischka, J., Mayer, S., Francisco, J., Gonçalves, D. C., et al. (2014). Molecular and biochemical characterization of caffeine synthase and purine alkaloid concentration in guarana fruit. *Phytochemistry* 105, 25–36. doi: 10.1016/j.phytochem.2014.04.018
- Shinbo, Y., Nakamura, Y., Altaf-Ul-Amin, M., Asahi, H., Kurokawa, K., Arita, M. et al. (2006). "KNAPSAck: a comprehensive species-metabolite relationship database," in *Plant Metabolomics*. Eds. K. R. A. Saito-Dixon and L. Willmitzer (Berlin, Heidelberg: Springer Berlin Heidelberg), 165–181. doi: 10.1007/3-540-29782-0_13
- Shulgin, A. T., and Perry, W. E. (2002). *The Simple Plant Isoquinolines*. Berkeley, Calif: Transform Press.
- Smith, T. A. (1977a). Phenethylamine and related compounds in plants. *Phytochemistry* 16, 9–18. doi: 10.1016/0031-9422(77)83004-5
- Smith, T. A. (1977b). Tryptamine and related compounds in plants. *Phytochemistry* 16, 171–175. doi: 10.1016/S0031-9422(00)86778-3
- Stenzel, O., Teuber, M., and Dräger, B. (2006). Putrescine *N*-methyltransferase in *Solanum tuberosum* L., a calystegine-forming plant. *Planta* 223, 200–212. doi: 10.1007/s00425-005-0077-z
- Sun, B., Wang, P., Wang, R., Li, Y., and Xu, S. (2018). Molecular cloning and characterization of a meta/para-*O*-methyltransferase from *Lycoris aurea*. *Int. J. Mol. Sci.* 19, 1911. doi: 10.3390/ijms19071911
- Suzuki, T., and Takahashi, E. (1975). Biosynthesis of caffeine by tea-leaf extracts. *Biochem. J.* 146, 87–96. doi: 10.1042/bj1460087
- Suzuki, T., and Takahashi, E. (1976). Caffeine biosynthesis in *Camellia sinensis*. *Phytochemistry* 15, 1235–1239. doi: 10.1016/0031-9422(76)85084-4
- Świderek, K., Tuñón, I., Williams, I. H., and Moliner, V. (2018). Insights on the origin of catalysis on glycine *N*-methyltransferase from computational modeling. *J. Am. Chem. Soc.* 140, 4327–4334. doi: 10.1021/jacs.7b13655
- Takeshita, N., Fujiwara, H., Mimura, H., Fitch, J. H., Yamada, Y., and Sato, F. (1995). Molecular cloning and characterization of *S*-adenosyl-L-methionine:scoulerine-9-*O*-methyltransferase from cultured cells of *Coptis japonica*. *Plant Cell Physiol.* 36, 29–36. doi: 10.1074/jbc.M106405200
- Tamura, K., Fujii, N., Morishige, T., Inui, T., and Sato, F. (2006). Overexpression of *Coptis japonica* norcoclaurine 6-*O*-methyltransferase overcomes the rate-limiting step in benzylisoquinoline alkaloid biosynthesis in cultured *Eschscholzia californica*. *Plant Cell Physiol.* 48, 252–262. doi: 10.1093/pcp/pcl062
- Taylor, J. S., and Raes, J. (2004). Duplication and divergence: the evolution of new genes and old ideas. *Annu. Rev. Genet.* 38, 615–643. doi: 10.1146/annurev.genet.38.072902.092831
- Teuber, M., Azemi, M. E., Namjooyan, F., Meier, A.-C., Wodak, A., Brandt, W., et al. (2007). Putrescine *N*-methyltransferases—a structure–function analysis. *Plant Mol. Biol.* 63, 787–801. doi: 10.1007/s11103-006-9126-7

- Thornton, J. W. (2004). Resurrecting ancient genes: experimental analysis of extinct molecules. *Nat. Rev. Genet.* 5, 366–375. doi: 10.1038/nrg1324
- Torres, M. A., Hoffarth, E., Eugenio, L., Savtchouk, J., Chen, X., Morris, J. S., et al. (2016). Structural and functional studies of pavine *N*-methyltransferase from *Thalictrum flavum* reveal novel insights into substrate recognition and catalytic mechanism. *J. Biol. Chem.* 291, 23403–23415. doi: 10.1074/jbc.M116.747261
- Trenchard, I. J., and Smolke, C. D. (2015). Engineering strategies for the fermentative production of plant alkaloids in yeast. *Metab. Eng.* 30, 96–104. doi: 10.1016/j.ymben.2015.05.001
- Uefuji, H. (2003). Molecular cloning and functional characterization of three distinct *N*-methyltransferases involved in the caffeine biosynthetic pathway in coffee plants. *Plant Physiol.* 132, 372–380. doi: 10.1104/pp.102.019679
- Wang, P., Hummel, E., Osterrieder, A., Meyer, A. J., Frigerio, L., Sparkes, I., et al. (2011). KMS1 and KMS2, two plant endoplasmic reticulum proteins involved in the early secretory pathway. *Plant J.* 66, 613–628. doi: 10.1111/j.1365-3113.2011.04522.x
- Waterman, P. G. (1998). “Chemical taxonomy of alkaloids,” in *Alkaloids: Biochemistry, Ecology, and Medicinal Applications*. Eds. M. F. Roberts and M. Wink (Boston, MA: Springer US), 87–107. doi: 10.1007/978-1-4757-2905-4_4
- Wei, C., Yang, H., Wang, S., Zhao, J., Liu, C., Gao, L., et al. (2018). Draft genome sequence of *Camellia sinensis* var. *sinensis* provides insights into the evolution of the tea genome and tea quality. *Proc. Natl. Acad. Sci.* 115, E4151–E4158. doi: 10.1073/pnas.1719622115
- Weng, J. K. (2014). The evolutionary paths towards complexity: a metabolic perspective. *New Phytol.* 201, 1141–1149. doi: 10.1111/nph.12416
- Wessjohann, L., Dippe, M., Tengg, M., and Gruber-Khadjawi, M. (2014). “Methyltransferases in biocatalysis,” in *Cascade Biocatalysis: Integrating Stereoselective and Environmentally Friendly Reactions*, (Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA), 393–426. doi: 10.1002/9783527682492.ch18
- Wink, M. (1984). *N*-Methylation of quinolizidine alkaloids: an *S*-adenosyl-L-methionine: cytosine *N*-methyltransferase from *Laburnum anagyroides* plants and cell cultures of *L. alpinum* and *Cytisus canariensis*. *Planta* 161, 339–344. doi: 10.1007/BF00398724
- Winzer, T., Gazda, V., He, Z., Kaminski, F., Kern, M., Larson, T. R., et al. (2012). A *Papaver somniferum* 10-gene cluster for synthesis of the anticancer alkaloid noscapine. *Science* 336, 1704–1708. doi: 10.1126/science.1220757
- Xia, E. H., Zhang, H. B., Sheng, J., Li, K., Zhang, Q. J., Kim, C., et al. (2017). The tea tree genome provides insights into tea flavor and independent evolution of caffeine biosynthesis. *Mol. Plant* 10, 866–877. doi: 10.1016/j.molp.2017.04.002
- Xu, Q., and Lin, M. (1999). Benzylisoquinoline alkaloids from *Gnetum parvifolium*. *J. Nat. Prod.* 62, 1025–1027. doi: 10.1021/np980472f
- Yoneyama, N., Morimoto, H., Ye, C.-X., Ashihara, H., Mizuno, K., and Kato, M. (2006). Substrate specificity of *N*-methyltransferase involved in purine alkaloids synthesis is dependent upon one amino acid residue of the enzyme. *Mol. Genet. Genomics* 275, 125–135. doi: 10.1007/s00438-005-0070-z
- Ziegler, J., and Facchini, P. J. (2008). Alkaloid biosynthesis: metabolism and trafficking. *Annu. Rev. Plant Biol.* 59, 735–769. doi: 10.1146/annurev.arplant.59.032607.092730
- Ziegler, J., Diaz-Chávez, M. L., Kramell, R., Ammer, C., and Kutchan, T. M. (2005). Comparative macroarray analysis of morphine containing *Papaver somniferum* and eight morphine free *Papaver* species identifies an *O*-methyltransferase involved in benzylisoquinoline biosynthesis. *Planta* 222, 458–471. doi: 10.1007/s00425-005-1550-4
- Zubaran, C. (2000). Ibogaine and noribogaine: comparing parent compound to metabolite. *CNS Drug Rev.* 6, 219–240. doi: 10.1111/j.1527-3458.2000.tb00149.x
- Zubieta, C. (2002). Structural basis for the modulation of lignin monomer methylation by caffeic acid/5-hydroxyferulic acid 3/5-*O*-methyltransferase. *Plant Cell Online* 14, 1265–1277. doi: 10.1105/tpc.001412
- Zubieta, C., He, X. Z., Dixon, R. A., and Noel, J. P. (2001). Structures of two natural product methyltransferases reveal the basis for substrate specificity in plant *O*-methyltransferases. *Nat. Struct. Biol.* 8, 271–279. doi: 10.1038/85029
- Zubieta, C., Ross, J. R., Koscheski, P., Yang, Y., Pichersky, E., and Noel, J. P. (2003). Structural basis for substrate recognition in the Salicylic Acid Carboxyl Methyltransferase family. *Plant Cell* 15, 1704–1716. doi: 10.1105/tpc.014548

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Morris and Facchini. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



OPEN ACCESS

Approved by:
Frontiers Editorial Office,
Frontiers Media SA, Switzerland

***Correspondence:**
Peter J. Facchini
pfacchin@ucalgary.ca

Specialty section:
This article was submitted to
Plant Metabolism
and Chemodiversity,
a section of the journal
Frontiers in Plant Science

Received: 15 October 2019

Accepted: 16 October 2019

Published: 08 November 2019

Citation:
Morris JS and Facchini PJ
(2019) Corrigendum: Molecular
Origins of Functional Diversity
in Benzyloquinoline Alkaloid
Methyltransferases.
Front. Plant Sci. 10:1436.
doi: 10.3389/fpls.2019.01436

Corrigendum: Molecular Origins of Functional Diversity in Benzyloquinoline Alkaloid Methyltransferases

Jeremy S. Morris and Peter J. Facchini*

Department of Biological Sciences, University of Calgary, Calgary, AB, Canada

Keywords: Benzyloquinoline, Alkaloid, Methyltransferase, specialized metabolism, molecular evolution

A Corrigendum On

Molecular Origins of Functional Diversity in Benzyloquinoline Alkaloid Methyltransferases
By Morris JS and Facchini PJ (2019). *Front. Plant Sci.* 10:1058. doi: 10.3389/fpls.2019.01058

In the original article, the author names were ordered incorrectly as “Peter J. Facchini and Jeremy S. Morris”. The correct order is “Jeremy S. Morris and Peter J. Facchini.”

The authors apologize for this error and state that this does not change the scientific conclusions of the article in any way. The original article has been updated.

Copyright © 2019 Morris and Facchini. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Biosynthesis and Emission of Stress-Induced Volatile Terpenes in Roots and Leaves of Switchgrass (*Panicum virgatum* L.)

Andrew Muchlinski^{1†}, Xinlu Chen^{2†}, John T. Lovell³, Tobias G. Köllner⁴, Kyle A. Pelot⁵, Philipp Zerbe⁵, Meredith Ruggiero¹, LeMar Callaway III¹, Suzanne Laliberte¹, Feng Chen^{2*} and Dorothea Tholl^{1*}

OPEN ACCESS

Edited by:

Kazuki Saito,
RIKEN Center for Sustainable
Resource Science (CSRS),
Japan

Reviewed by:

Takao Koeduka,
Yamaguchi University,
Japan
Yasuhiro Higashi,
RIKEN Yokohama,
Japan

*Correspondence:

Feng Chen
fengc@utk.edu
Dorothea Tholl
tholl@vt.edu

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Plant Metabolism and
Chemodiversity,
a section of the journal
Frontiers in Plant Science

Received: 21 April 2019

Accepted: 21 August 2019

Published: 19 September 2019

Citation:

Muchlinski A, Chen X, Lovell J,
Köllner TG, Pelot KA, Zerbe P,
Ruggiero M, Callaway L III,
Laliberte S, Chen F and Tholl D
(2019) Biosynthesis and Emission
of Stress-Induced Volatile Terpenes
in Roots and Leaves of Switchgrass
(*Panicum virgatum* L.).
Front. Plant Sci. 10:1144.
doi: 10.3389/fpls.2019.01144

¹ Department of Biological Sciences, Virginia Tech, Blacksburg, VA, United States, ² Department of Plant Sciences, University of Tennessee, Knoxville, TN, United States, ³ Genome Sequencing Center, Hudson Alpha Institute for Biotechnology, Huntsville, AL, United States, ⁴ Department of Biochemistry, Max Planck Institute for Chemical Ecology, Jena, Germany, ⁵ Department of Plant Biology, University of California, Davis, Davis, CA, United States

Switchgrass (*Panicum virgatum* L.), a perennial C4 grass, represents an important species in natural and anthropogenic grasslands of North America. Its resilience to abiotic and biotic stress has made switchgrass a preferred bioenergy crop. However, little is known about the mechanisms of resistance of switchgrass against pathogens and herbivores. Volatile compounds such as terpenes have important activities in plant direct and indirect defense. Here, we show that switchgrass leaves emit blends of monoterpenes and sesquiterpenes upon feeding by the generalist insect herbivore *Spodoptera frugiperda* (fall armyworm) and in a systemic response to the treatment of roots with defense hormones. Belowground application of methyl jasmonate also induced the release of volatile terpenes from roots. To correlate the emission of terpenes with the expression and activity of their corresponding biosynthetic genes, we identified a gene family of 44 monoterpene and sesquiterpene synthases (mono- and sesqui-TPSs) of the type-a, type-b, type-g, and type-e subfamilies, of which 32 TPSs were found to be functionally active *in vitro*. The TPS genes are distributed over the K and N subgenomes with clusters occurring on several chromosomes. Synteny analysis revealed syntenic networks for approximately 30–40% of the switchgrass TPS genes in the genomes of *Panicum hallii*, *Setaria italica*, and *Sorghum bicolor*, suggesting shared TPS ancestry in the common progenitor of these grass lineages. Eighteen switchgrass TPS genes were substantially induced upon insect and hormone treatment and the enzymatic products of nine of these genes correlated with compounds of the induced volatile blends. In accordance with the emission of volatiles, TPS gene expression was induced systemically in response to belowground treatment, whereas this response was not observed upon aboveground feeding of *S. frugiperda*. Our results demonstrate complex above and belowground responses of induced volatile terpene metabolism in switchgrass and provide a framework for more detailed investigations of the function of terpenes in stress resistance in this monocot crop.

Keywords: switchgrass, terpene synthase, volatile, herbivory, defense

INTRODUCTION

Switchgrass (*Panicum virgatum* L., Poaceae) is a native warm-season C4 perennial grass common to natural and anthropogenic grasslands in North America. Characteristic of the Tallgrass Prairie, switchgrass is considered an important species for sustaining natural prairie biodiversity (Sanderson et al., 2006). Used mostly for forage since the 1950s, more intensive breeding of switchgrass began over 20 years ago to develop the species as an herbaceous model species for biofuel feedstock development (Casler et al., 2011). Major advantages for cultivating switchgrass are its resilience to extreme weather conditions, capability of growing on marginal soils, and a high cellulosic content (Vogel, 2004). Switchgrass also exhibits considerable resistance to pests and diseases (Parrish and Fike, 2005). With an increase in cultivation, growing interest has focused on elucidating the resistance mechanisms of switchgrass as well as engineering more resistant varieties. However, surprisingly little is still known about the modes of pathogen and pest defense in this species.

Plants deploy a biosynthetic and structurally diverse mosaic of specialized or secondary metabolites for chemical defense (Dudareva et al., 2004). Terpenes constitute the majority of such metabolites with important defensive activities. For instance, nonvolatile triterpenes are potent growth inhibitors of fungal pathogens (Osborn, 1996). By contrast, low molecular weight 10-carbon monoterpenes and 15-carbon sesquiterpenes are emitted by plants as volatile compounds and serve important roles in direct defenses against pathogens and herbivores or function indirectly by the attraction of parasitoids or intra- and interplant priming (Turlings et al., 1990; Dicke, 1994; Kost and Heil, 2006; Köllner et al., 2008a; Huang et al., 2012; Vaughan et al., 2013; Erb et al., 2015).

The formation of terpenes in plants is catalyzed by enzymes of the terpene synthase superfamily (TPSs). TPS enzymes convert 10- and 15-carbon *cis*- or *trans*-isoprenyl diphosphates such as geranyl diphosphate (GDP), neryl diphosphate, farnesyl diphosphate [(*E,E*)-FDP or (*Z,Z*)-FDP] into monoterpenes or sesquiterpenes, respectively (Tholl and Lee, 2011). TPS genes often undergo species specific divergence and duplications resulting in terpene metabolic plasticity and adaptations (Pichersky and Gang, 2000). The structural diversity and biosynthetic evolution of terpene secondary metabolites have been studied extensively in crops including grasses such as maize, rice, and sorghum (Chen et al., 2011; Boutanaev et al., 2015; Block et al., 2019). Terpene-related defenses have been well described in these monocot crops and reveal diverse chemical mechanisms for resistance against above- and belowground stressors. For example, the sesquiterpene (*E*)- β -caryophyllene, one of the major volatile organic compounds (VOCs) released by maize leaves and roots, is involved in indirect defense by attracting parasitoids of herbivores and entomopathogenic nematodes (Turlings et al., 1990; Rasmann et al., 2005; Köllner et al., 2008a). Monoterpenes have also been implicated in defensive roles; for example, linalool confers resistance against rice bacterial blight caused by *Xanthomonas oryzae* (Taniguchi et al., 2014). More recently, a rice (*S*)-limonene synthase (*OsTPS19*) was shown to be involved in direct defense against the blast fungus *Magnaporthe oryzae* (Chen et al., 2018).

In contrast to these findings in highly domesticated grasses, the biosynthesis and dynamics of terpenes in switchgrass have not been fully investigated, in part because of its complex genetic background. Lowland ecotypes are allotetraploid ($2n = 4x = 36$), while upland cultivars are frequently octoploid ($2n = 8x = 72$). Recent transcriptional analysis of defense responses to green bug herbivory (*Schizaphis graminum*, Aphididae) in switchgrass leaves revealed a global transcriptional remodeling resulting in increased reactive oxygen species production and upregulation of genes with predicted terpene synthase function (Donze-Reiner et al., 2017). Moreover, the presence of a few triterpene saponins (C_{30}) (Lee et al., 2009) and the synthesis of diterpenes (C_{20}) related to abiotic stress have been described (Pelot et al., 2018). However, no prior studies have investigated the formation and function of volatile terpenes in this grass. Therefore, we sought to identify and characterize TPS genes from the switchgrass genome and correlate stress-induced terpene synthases with compound production in roots and leaves. Particular focus was placed on TPSs that were readily inducible when challenged with a generalist herbivore and the defense-related phytohormones methyl jasmonate (MeJA) and salicylic acid (SA) with the future goal to investigate these genes for their broad defensive functions against pathogens and herbivores. Results from this study provide further insight into the genetic organization of terpene metabolism in switchgrass and illustrate the metabolic potential of terpene-related defenses in perennial polyploid grasses.

MATERIALS AND METHODS

Plant Materials

Seeds from the lowland allotetraploid switchgrass cv. Alamo were purchased from Bamert Seed Company (Muleshoe, TX) and used throughout this study. The seeds were sowed into potting substrate in 200-ml aluminum cans or 2.5" pots and grown for 5 weeks at 26°C (16 h day) and 24°C (8 h night) in a Percival growth chamber. After germination, 15 seedlings were selected in each can or pot and grown for 5 weeks.

Plant Treatments

Five-week-old seedlings were treated with larvae of *S. frugiperda* as described by Zhuang et al. (2012) with some modifications. Cans with 15 seedlings were each placed into a collection chamber, and 10 second instar larvae were released inside the chamber for overnight feeding. For treatment with MeJA and SA (Sigma-Aldrich), 25 ml of MeJA (0.1, 1, and 5 mM) or SA (5 mM) dissolved in ethanol were added per can or pot as a soil drench and left for 24 h, respectively. For physical wounding, a surgical scalpel was used to wound leaves and stems. Untreated plants and mock-treated plants (ethanol only) were used as controls. Three replicates were performed for each treatment.

Volatile Collection and Identification

Volatiles emitted from leaves of the treated switchgrass and control plants placed in glass chambers were collected with an open headspace sampling system (Analytical Research Systems,

Gainesville, FL, USA) in the light from 9:00 AM to 1:00 PM. Fall armyworm (FAW) larvae were removed before volatile trapping. The volatiles were collected with volatile collection traps (Porapak-Q, <http://www.volatilecollectiontrap.com/>) and eluted with 100 μ l methylene chloride containing 0.003% nonyl acetate (v/v). The collected volatiles were analyzed on a Shimadzu 17A gas chromatograph coupled to a Shimadzu QP5050A (<http://www.shimadzu.com>). Statistical analysis of leaf volatile data was done in R (v3.5.0) using ANOVA and *post-hoc* Tukey–Kramer honestly significant difference comparisons where $\alpha \leq 0.05$.

Root volatiles were analyzed by automated solid-phase microextraction (SPME, AOC-5000 Shimadzu) through adsorption in the headspace with a 100- μ m polydimethylsiloxane (Supelco) SPME fiber and thermal desorption for gas chromatography–mass spectrometry (GC-MS) analysis. Root tissue (1 g fresh weight) was detached from plants and placed in a 20-ml screw-capped vial containing 2 ml distilled water and 20 ng of the volatile internal standard 1-bromodecane. The SPME fiber was placed into the headspace of the vial and incubated for 30 min at room temperature for volatile collection. Collected volatiles were thermally desorbed for 4 min and analyzed using a gas chromatograph (240°C injector port) coupled with a quadrupole mass spectrometer (GC-MS-QP2010S, Shimadzu). Extracts were separated with a 2:1 split on a 30 m \times 0.25 mm i.d. \times 0.25 μ m film thickness Zebron capillary column (Phenomenex) using helium as the carrier gas (1.4 ml min^{−1} flowrate) and a temperature gradient of 5°C min^{−1} from 40°C (hold 2 min) to 220°C. Compound identification was based on similarity to library matches (NIST, Wiley), authentic standards (Sigma-Aldrich, (*E*)- β -caryophyllene, germacrene-D), and comparison to Opopanax essential oil (Floracopeia, δ -cadinene, α -humulene). Relative abundance was determined by normalization of the analyte peak area to the peak area of the internal standard and dividing by gram fresh weight.

Identification of TPS Genes From the Switchgrass Genome and Phylogeny Reconstruction

Putative switchgrass TPS genes were retrieved from Phytozome (www.phytozome.jgi.doe.gov) through an annotation-based keyword search of genome versions v.1 and v.4. In addition, RNA-seq data kindly provided by the Noble Foundation (<https://www.noble.org>) for above- and belowground tissues were assembled *de novo* using Trinity (Grabherr et al., 2011). Assembled transcriptomes were queried with a representative switchgrass TPS sequence (*PvTPS01*) using the National Center for Biotechnology Information's TBLASTX. Resulting BLAST hits were manually curated for putative functionality based on length and presence of the conserved aspartate rich motif (DDxxD) necessary for ionization of the prenilyldiphosphate substrate (Class I TPSs). Class I and II diterpene synthases identified in this study were not further pursued based on previous reporting by Pelot et al. (2018). Gene models were refined further by comparing transcripts to genome sequences available in Phytozome. Putative N-terminal plastidic transit peptides were predicted using

multiple sequence alignments and analysis of each sequence with the transit peptide prediction software ChloroP (Emanuelsson et al., 1999). Phylogeny reconstruction was based on protein sequence alignments, which were performed using MAFFT (Katoh et al., 2002). Maximum likelihood trees were then built from MAFFT alignments using PhyML (Guindon et al., 2010) with 500 bootstrap replicates as previously described (Pelot et al., 2018). Final phylogeny annotation and design were performed in Interactive Tree of Life (Letunic and Bork, 2007). Heat map analysis was based on publicly available expression data at <http://www.phytozome.net/> following previously described methods (Pelot et al., 2018).

Synteny Analysis and Identification of Orthologous TPS Genes

P. virgatum (v4.1), *Setaria italica* (v2.2), *Sorghum bicolor* (v3.1), and *P. hallii* var. *hallii* (v2.1) genome annotations were downloaded from phytozome (phytozome.jgi.doe.gov). Syntenic blocks were generated following Lovell et al. (2018) via the GENESPACE pipeline. Orthofinder was run on synteny-constrained BLASTp results to build orthologous gene networks.

Gene Expression Analysis

Total RNA was isolated from switchgrass leaves and roots using the RNeasy Plant Mini Kit according to the manufacturer's protocol (<http://www.qiagen.com>). Complementary DNA was synthesized using the GE Healthcare first-strand synthesis kit according to the manufacturer's protocol (<http://www.gelifesciences.com>). Gene expression analysis was carried out using quantitative reverse transcription PCR (RT-PCR), which was described previously (Chen et al., 2018). Sequences of primers used for RT-qPCR are listed in **Supplementary Table 3**.

Protein Expression in *E. coli* and Terpene Synthase Activity Assay

Full-length and truncated genes (predicted transit peptide removed) were synthesized and cloned (*Nde*I) into the pET-28b(+) prokaryotic expression vector. Constructs were transformed into *Escherichia coli* BL21-CodonPlus(DE3) cells (Stratagene) and grown at 37°C in 100 ml Luria–Bertani media supplemented with 50 μ M kanamycin until an optical density at 600 nm (OD₆₀₀) of 0.5–0.7. Protein production was then induced with 0.5 mM isopropyl 1-thio- β -D-galactopyranoside, and cells were incubated with shaking at 18°C for 16 h. Recombinant protein extraction and partial purification were performed as described by Tholl et al. (2005), with the modification that N-terminal His-tags were implemented for partial purification. Enzyme reactions (125 μ l total volume) were prepared in a 10-ml screw cap vial (Supelco) by combining partially purified protein with 20 mM MgCl₂ and 60 μ M commercially available prenilyl diphosphate substrates GDP and (*E,E*)-FDP (Echelon Biosciences). Assay mixtures were incubated for 5 min at 30°C in the presence of a 100- μ m polydimethylsiloxane fiber (Supelco). Collected volatiles were thermally desorbed for 4 min and analyzed using a gas

chromatograph (240°C injector port) coupled with a quadrupole mass spectrometer (GC-MS-QP2010S, Shimadzu). Extracts were separated with a 5:1 split under the same conditions described above. Compound identification, in addition to those compounds described above, was based on similarity to library matches (NIST, Wiley, copaene, cycloisosativene, β -elemene, α -patchoulene, α -selinene, valencene), authentic standards (Sigma-Aldrich, borneol, 1,8-cineole, geraniol, limonene, linalool, α -pinene, sabinene, α -terpineol, α -terpinolene), and comparison to Opopanax oil [Floracopeia, β -bisabolene, (*E*)- γ -bisabolene, γ -curcumene, (*E*)- β -farnesene, sabinene, α -santalene].

RESULTS

Emission of Volatile Terpenes From Leaves in Response to Insect Feeding

To assess whether switchgrass leaves emit volatile compounds upon aboveground herbivory, emissions from switchgrass plants (cv. Alamo) damaged by larvae of *S. frugiperda* (FAW) were collected by open headspace sampling and analyzed by GC-MS. We found that FAW treatment induced the emission of nine terpene compounds, which were not detected in plants that only received physical wounding or remained untreated (Figure 1 and Supplementary Table 1). Among the released compounds, the sesquiterpenes (*E*)- β -caryophyllene and (*E*)- β -farnesene were strongly induced by herbivore damage

accounting for ~17 and ~26%, respectively, of the total volatile organic compound emission (Figure 1 and Supplementary Table 1). Emission rates of (*E*)- β -caryophyllene were ~500 ng/h g FW. Additional major compounds induced by FAW included the monoterpene (*E*)- β -ocimene, the homoterpene (*E*)-DMNT, and the sesquiterpenes β -elemene, α -bergamotene, α -humulene, and β -copaene (Figure 1 and Supplementary Table 1).

Emission of Volatile Terpenes From Roots and Leaves Upon Belowground Treatment With Methyl Jasmonate or Salicylic Acid

We further determined whether emissions of volatile compounds from switchgrass roots could be induced by root treatment with phytohormones mimicking herbivory or pathogen infection. Different concentrations of MeJA were tested (0.1, 1, and 5 mM) by watering plants directly with each solution. Because of the volatility of MeJA, we expected that the compound diffused further into the substrate at a lower concentration. Volatiles were collected from detached roots using SPME and analyzed by GC-MS. Concentrations of 1 and 5 mM MeJA caused a similar relative release of sesquiterpene compounds from the root tissue (shown for 5 mM treatment; Figure 2), while no volatiles were induced upon treatment with 0.1 mM MeJA. Of the seven identified compounds, (*E*)- β -caryophyllene was the most abundant (~43% of total), while cycloisosativene, β -elemene, α -humulene, α -selinene, germacrene D, and δ -cadinene were

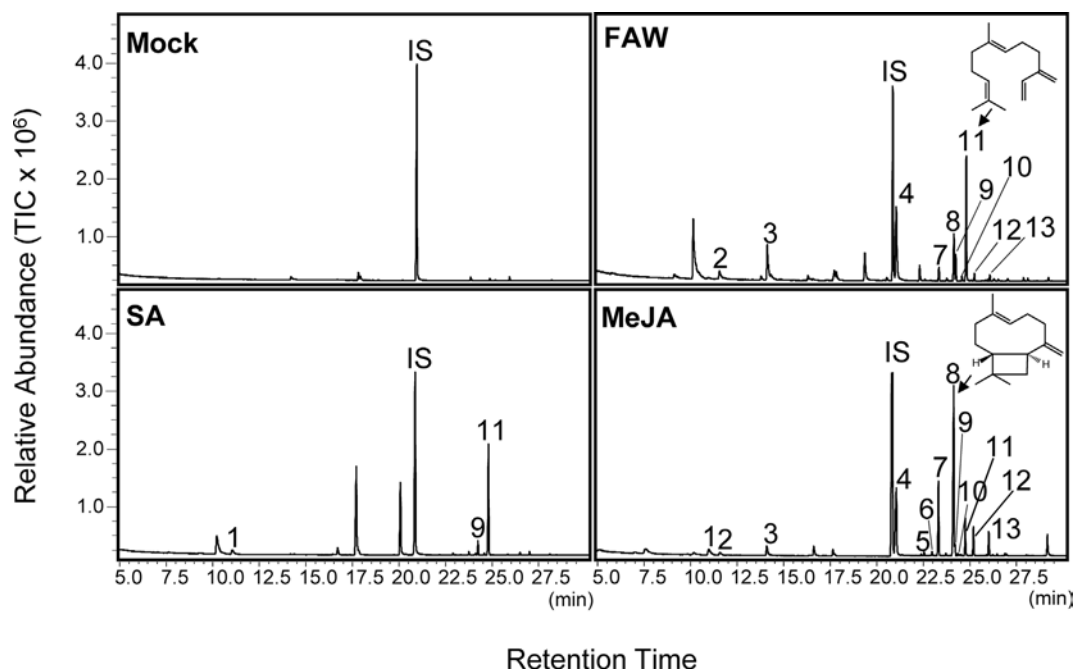


FIGURE 1 | Gas chromatography–mass spectrometry (GC-MS) analysis of volatile emission from leaves of switchgrass (cv. Alamo) treated with fall armyworm (FAW), salicylic acid (SA), or methyl jasmonate (MeJA). Individual chromatograms are representative of samples analyzed in biological triplicates. Compound identification was based on similarity to library matches (NIST, Wiley) and comparisons to authentic standards. 1: limonene; 2: (*E*)- β -ocimene; 3: (*E*)-DMNT; 4: indole; 5: α -ylangene; 6: elemene isomer; 7: β -elemene; 8: (*E*)- β -caryophyllene; 9: α -bergamotene; 10: unidentified sesquiterpene; 11: (*E*)- β -farnesene; 12: α -humulene; and 13: β -copaene. IS: internal standard nonyl acetate.

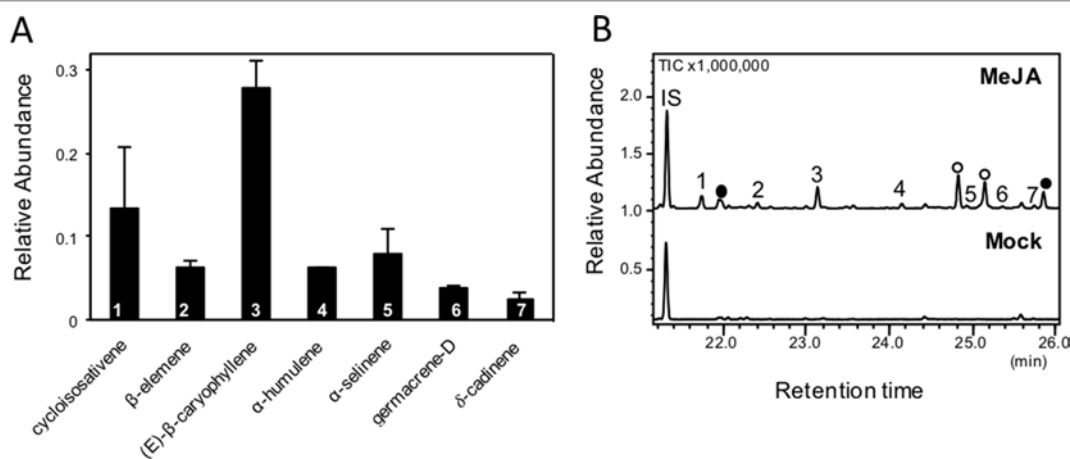


FIGURE 2 | Solid-phase microextraction–gas chromatography–mass spectrometry (SPME–GC–MS) analysis of root sesquiterpene emission following methyl jasmonate (MeJA) treatment. Volatile compounds were analyzed in triplicate from detached pooled root material of 5-week-old plants. No sesquiterpenes were identified in the mock treatment. **(A)** Relative abundance of the detected sesquiterpenes. Samples were normalized to the internal standard and gram fresh weight. **(B)** Representative gas chromatograms of emitted volatiles in MeJA- and mock-treated roots. IS: internal standard 1-bromodecane; shaded circles indicate putative terpene compounds, and open circles indicate putative fatty acid derivatives. Compound identification was based on similarity to library matches (NIST, Wiley) and authentic standards.

present at low levels. We also applied SA at a concentration of 5 mM; however, no release of sesquiterpenes was observed from root tissue. We further found two monoterpenoids, camphor and borneol, to be released from root tissue of untreated plants. Emissions of these compounds were reduced by MeJA and SA treatments, although this was not found to be statistically significant based on comparisons of the means (ANOVA, $p > 0.05$, **Supplementary Figure S1**).

We also tested whether a drench with MeJA and SA at 5 mM could induce volatile emissions in aboveground tissues. Treatment with MeJA strongly induced volatile emission from leaves compared to mock controls, with 13 compounds identified (**Figure 1** and **Supplementary Table 1**). Major induced compounds were (*E*)- β -caryophyllene and β -elemene accounting for ~38 and ~17%, respectively, of total volatile emissions (**Figure 1** and **Supplementary Table 1**). Emission rates of (*E*)- β -caryophyllene were approximately 1,500 ng/h g FW. Other minor compounds included limonene, (*E*)- β -ocimene, (*E*)-DMNT, α -ylangene, α -bergamotene, α -humulene, (*E*)- β -farnesene, and β -copaene (**Figure 1** and **Supplementary Table 1**). Two additional putative sesquiterpenes were also emitted; however, these compounds could not be further identified based on available standards. Treatment with SA induced the emission of four terpene compounds with (*E*)- β -farnesene accounting for 83% of total emissions. Trace amounts of limonene, (*E*)- β -caryophyllene, and α -bergamotene were detected, which were not observed in the untreated controls (**Figure 1** and **Supplementary Table 1**).

Genome-Wide Identification of Putative Terpene Synthases in Switchgrass

Based on the inducible emission of diverse volatile terpenes from switchgrass roots and leaves, we sought to identify the TPS genes

responsible for their formation. Following a genome-wide search of the switchgrass draft genome v.1, we originally identified 144 putative TPS gene models. Of these putative gene models, 108 were confirmed in the draft genome v.4, with 74 putative full-length mono-, sesqui-, and di-TPS genes identified. Manual sequence curation through multiple sequence alignments and comparison to genomic and transcriptomic data resulted in the identification of 44 putative full-length mono- and sesqui-TPS genes (**Table 1**, **Supplementary Table 2**, **Supplementary Figure S2**). Identified di-TPS genes (30 in total) were previously reported and therefore not included in this study (Pelot et al., 2018).

Alignment and phylogenetic analysis of amino acid sequences from the mono- and sesqui-TPSs together with select TPSs from maize, rice, sorghum, tomato, and snapdragon showed that 35 members belong to the TPS type-a clade (**Figure 3**). In addition, five proteins aligned to the TPS-g subfamily and three clustered in the TPS-b subfamily. Only PvTPS15 (TPS-e) was predicted to be involved in volatile formation outside of the TPS-a, TPS-b, and TPS-g subfamilies. Like in other plant TPS proteins, switchgrass TPSs of the TPS-a, TPS-b, TPS-e, and TPS-g subfamilies carry the conserved aspartate-rich “DDXXD” motif and the less conserved “NSE/DTE” motif in the C-terminal α -domain (Chen et al., 2011).

When we examined the relative chromosomal position of the identified TPS genes, we found that 22 genes are distributed across the nine chromosomes in the switchgrass subgenome K with highest abundance of genes occurring on chromosomes 1K, 6K, and 9K (**Figure 4** and **Table 1**). In subgenome N, we identified the relative location of 20 genes with highest abundance on chromosomes 1N, 6N, and 9N (**Figure 4** and **Table 1**). Several genes are positioned in loose gene clusters throughout the genome (**Figure 4** and **Table 1**). The relative positions of PvTPS02 and PvTPS07 could not be determined based on incomplete genomic data (**Table 1**).

TABLE 1 | Identified mono- and sesqui-terpene synthase (mono- and sesqui-TPS) gene models in the switchgrass (cv. "AP13") genome in the order of chromosomal localization.

Locus ID	Designation	Protein Length	Genomic coordinates	Sub-family
Pavir.1KG024200	PvTPS28	520	Chr01K:2898821.2904002	tps-g
Pavir.1KG026700	PvTPS13	557	Chr01K:3056228.3059326	tps-g
Pavir.1KG213700	PvTPS83	607	Chr01K:30203541.30207022	tps-a
Pavir.1KG250000	PvTPS03	612	Chr01K:41080529.41082963	tps-a
Pavir.1KG359700	PvTPS33	574	Chr01K:60696889.60702775	tps-a
Pavir.1NG048300	PvTPS106	551	Chr01N:2491525.2496085	tps-g
Pavir.1NG047800	PvTPS27	552	Chr01N:2543130.2546026	tps-g
Pavir.1NG173900	PvTPS04	607	Chr01N:30384431.30393731	tps-a
Pavir.1NG245800	PvTPS26	583	Chr01N:56485952.56491693	tps-a
Pavir.2KG086400	PvTPS01	548	Chr02K:11130096.11132928	tps-a
Pavir.2KG150300	PvTPS14	552	Chr02K:19642302.19645362	tps-a
Pavir.2KG163100	PvTPS11	554	Chr02K:22460036.22464132	tps-a
Pavir.2NG122900	PvTPS85	515	Chr02N:18521205.18525163	tps-a
Pavir.2NG123000	PvTPS19	555	Chr02N:18533973.18537033	tps-a
Pavir.3KG098200	PvTPS36	592	Chr03K:8855955.8859293	tps-a
Pavir.3KG098400	PvTPS81	621	Chr03K:8870023.8873401	tps-a
Pavir.3KG102800	PvTPS08	616	Chr03K:9179806.9183931	tps-a
Pavir.5KG204300	PvTPS79	567	Chr05K:26556249.26559391	tps-a
Pavir.5KG258500	PvTPS09	491	Chr05K:38083739.38087664	tps-a
Pavir.6KG047500	PvTPS109	582	Chr06K:6526582.6531758	tps-a
Pavir.6KG066200	PvTPS17	549	Chr06K:9019105.9024173	tps-a
Pavir.6KG122500	PvTPS69	549	Chr06K:16365911.16369794	tps-a
Pavir.6KG141400	PvTPS10	259	Chr06K:23665655.23669023	tps-a
Pavir.6NG075100	PvTPS18	545	Chr06N:14008882.14012062	tps-a
Pavir.6NG075300	PvTPS16	626	Chr06N:14020511.14024883	tps-a
Pavir.6NG075700	PvTPS20	549	Chr06N:14043871.14051763	tps-a
Pavir.6NG192300	PvTPS94	545	Chr06N:36030001.36038500	tps-a
Pavir.6NG135600	PvTPS06	544	Chr06N:58557205.58561380	tps-a
Pavir.7NG405000	PvTPS15	618	Chr07N:63119161.63123381	tps-e
Pavir.8KG329300	PvTPS62	597	Chr08K:67429955.67433773	tps-a
Pavir.8NG101200	PvTPS101	549	Chr08N:15863965.15870973	tps-a
Pavir.9KG112900	PvTPS12	579	Chr09K:10384938.10389576	tps-b
Pavir.9KG454600	PvTPS56	570	Chr09K:66968728.66972114	tps-a
Pavir.9KG443500	PvTPS54	576	Chr09K:67815040.67822909	tps-a
Pavir.9KG442600	PvTPS55	552	Chr09K:67885565.67896951	tps-a
Pavir.9KG486800	PvTPS73	546	Chr09K:72969142.72975315	tps-b
Pavir.9NG517400	PvTPS50	551	Chr09N:86982070.86991194	tps-a
Pavir.9NG529400	PvTPS71	549	Chr09N:88336547.88343612	tps-a
Pavir.9NG574000	PvTPS05	560	Chr09N:92237810.92240737	tps-a
Pavir.9NG595500	PvTPS53	606	Chr09N:94252282.94256051	tps-a
Pavir.9NG595900	PvTPS104	467	Chr09N:94271551.94277158	tps-a
Pavir.9NG693900	PvTPS52	601	Chr09N:99983657.99989362	tps-b
Pavir.J27731	PvTPS07	502	contig312548:164.1302	tps-a
Pavir.J691000	PvTPS02	515	scaffold_732:11619.13941	tps-g

Genomic coordinates were determined based on draft genome data available in Phytozome (<https://phytozome.jgi.doe.gov/>).

Investigation of syntenic orthologous genes between the two switchgrass subgenomes identified networks between 8 genes on subgenome K and 10 genes on subgenome N (including one putative mono- or di-TPS, 3NG211100, and two putative di-TPSs, 3KG400900 and 3NG171200) (Figure 4, Supplementary Figure S3, Supplementary Table 2). Comparisons between the genomes of switchgrass and sorghum showed that 13 switchgrass TPS loci have syntenic orthologs on 6 of the 10 sorghum chromosomes (Figure 4, Supplementary Figure S3, Supplementary Table 2). Several of these switchgrass TPS genes also occur in syntenic gene networks with genomes of the more closely related grasses *Setaria italica* and *Panicum*

hallii suggesting conserved genomic regions in TPS gene evolution in these species.

Biochemical Characterization of Monoterpene and Sesquiterpene Synthases From Switchgrass

To determine the *in vitro* function of the 44 identified TPS genes, open reading frames were synthesized and cloned into the bacterial pET28b expression vector. The recombinant proteins were expressed in *E. coli* and protein lysates tested for TPS activity with GDP and (*E,E*)-FDP as substrates. We expected many TPSs

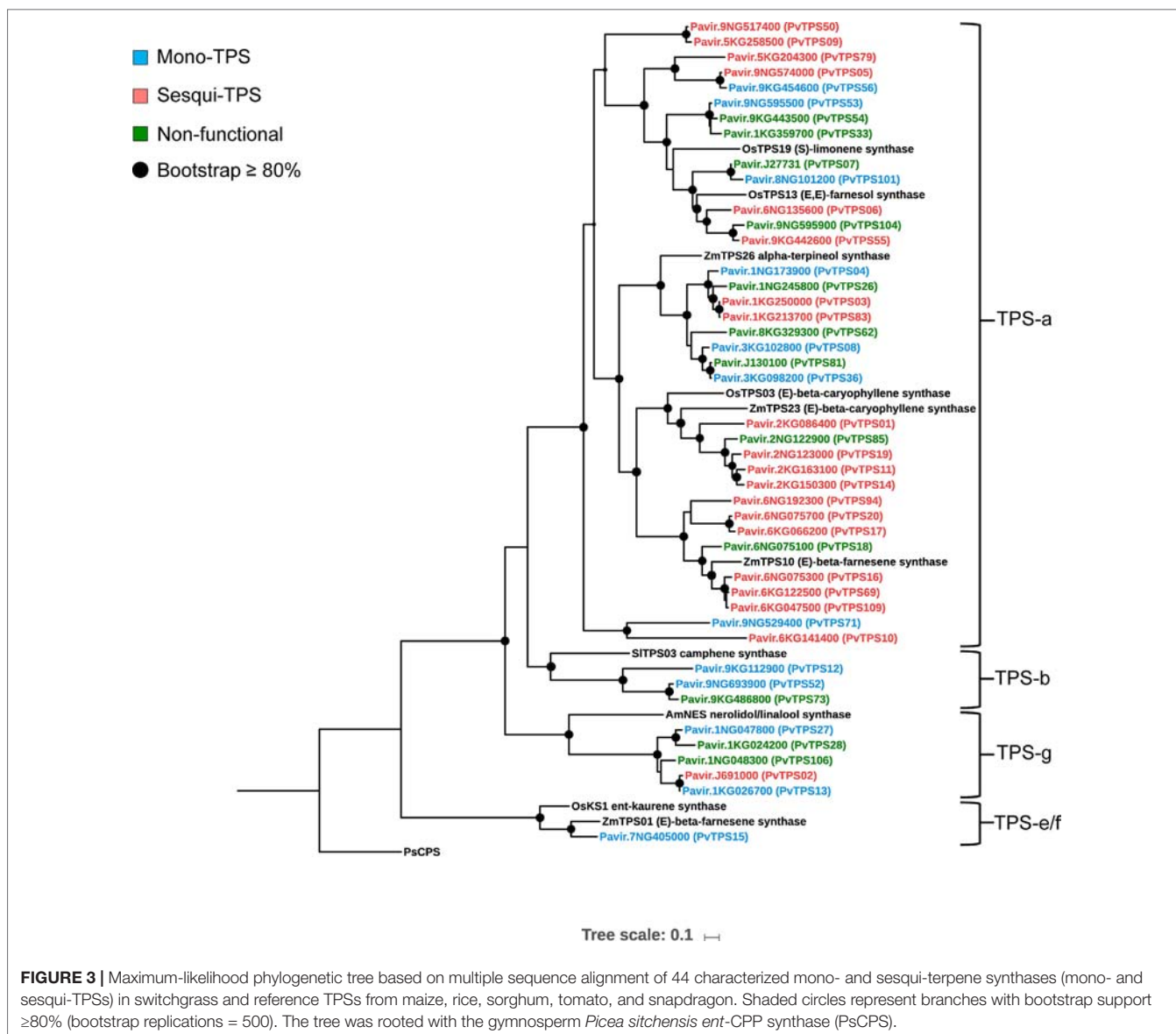


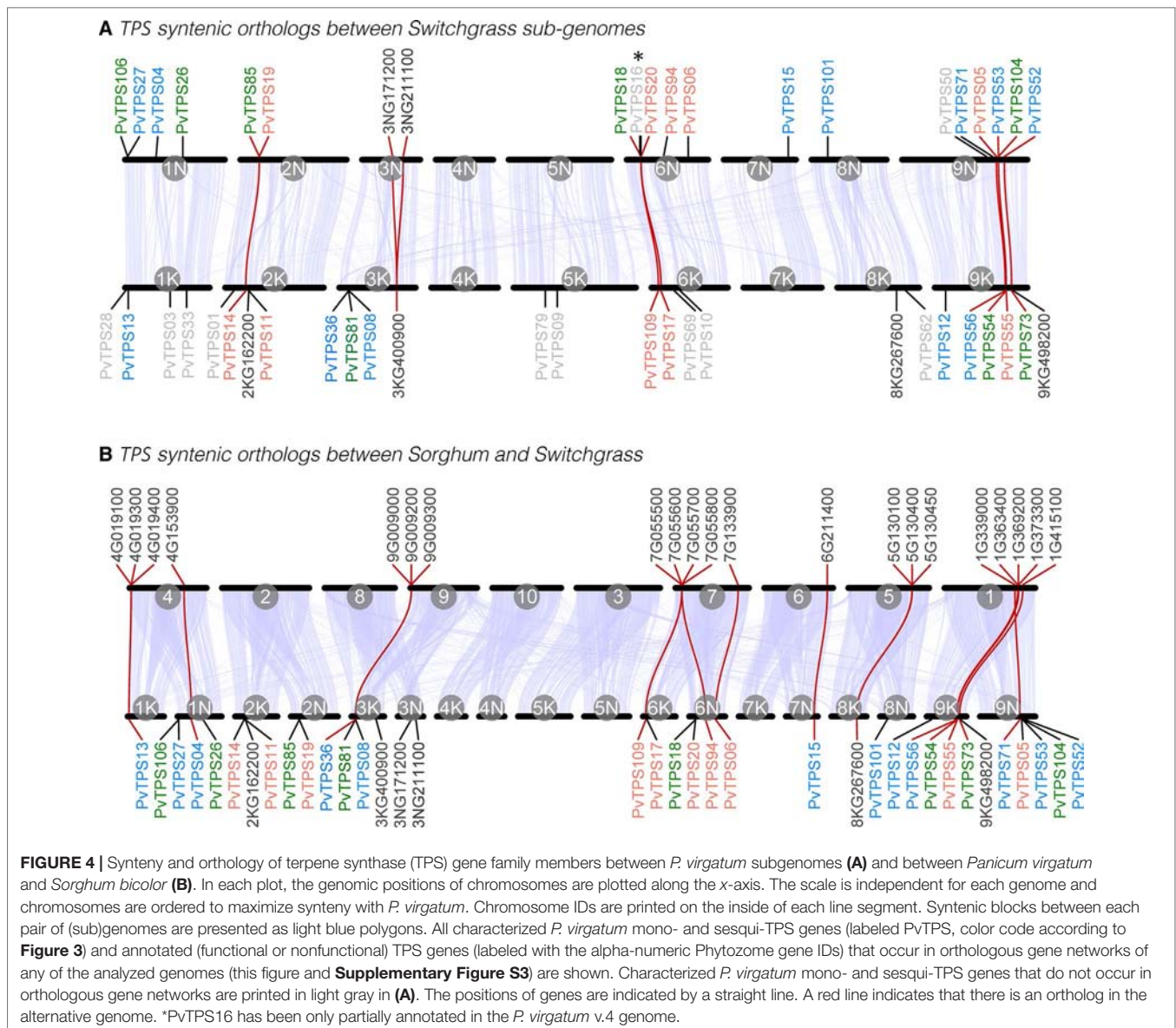
FIGURE 3 | Maximum-likelihood phylogenetic tree based on multiple sequence alignment of 44 characterized mono- and sesqui-terpene synthases (mono- and sesqui-TPSs) in switchgrass and reference TPSs from maize, rice, sorghum, tomato, and snapdragon. Shaded circles represent branches with bootstrap support $\geq 80\%$ (bootstrap replications = 500). The tree was rooted with the gymnosperm *Picea sitchensis* ent-CPP synthase (PsCPS).

in the subfamily-a (**Figure 3**) to function as sesqui-TPSs. Indeed, 19 recombinant TPS proteins in this family produced one or more sesquiterpene olefins, among them (*E*)- β -caryophyllene, (*E*)- β -farnesene, and other common plant sesquiterpenes (**Figures 3 and 5**). All of these proteins except PvTPS83 did not carry a plastidial transit peptide, indicating that they are likely to function in the cytosolic compartment. PvTPS02 was the only TPS protein found in the g-subfamily to exhibit sesquiterpene synthase activity *in vitro*. However, since a plastidial targeting sequence typical of subtype-g TPSs has been predicted for this protein, its function as a sesqui-TPS *in vivo* might be limited.

Twelve TPS proteins distributed over the TPS-a, TPS-b, TPS-g, and TPS-e subfamilies functioned as monoterpene synthases *in vitro* (**Figures 3 and 6**). PvTPS04 produced a mixture of monoterpenes from GDP with α -terpinolene and borneol as major products (**Figure 6**). PvTPS36 and PvTPS56 converted

GDP into multiple volatile products with predominantly limonene and α -terpineol as the major products, respectively (**Figure 6**). The remaining enzymes produced either linalool (PvTPS12, PvTPS13, PvTPS15, PvTPS27, PvTPS52, and PvTPS71) or geraniol (PvTPS53 and PvTPS101) (**Figure 6**). PvTPS13 and PvTPS15 also converted (*E,E*)-FDP into nerolidol (**Supplementary Figure S4**); however, this activity might be limited *in vivo* because of the predicted plastidial localization of these proteins. On the contrary, no plastidial transit peptides were predicted for PvTPS12, PvTPS56, PvTPS71, and PvTPS101, which questions their function as monoterpene synthases *in vivo*.

Only trace amounts of compounds were detected for recombinant proteins encoded by PvTPS07, PvTPS62, PvTPS81, and PvTPS106. In addition, no substantial enzymatic activity was found for eight proteins (PvTPS18, PvTPS26, PvTPS28, PvTPS33, PvTPS54, PvTPS73, PvTPS85, and PvTPS104), which is in accordance with



the presence of several deletions and/or insertions in the open reading frames of the corresponding genes (Supplementary Figure S2). Sequence truncations were furthermore found at the N- and C-terminus of the functionally active PvTPS09 and PvTPS02 proteins, respectively (Supplementary Figure S2).

Expression Analysis of PvTPSs in Different Tissues and Upon Treatment With FAW, MeJA, and SA

Global expression patterns for all 44 TPS genes were analyzed by hierarchical cluster analysis based on publicly available data (<https://phytozome.jgi.doe.gov/>). We found specific patterns of transcript abundance in vascular tissue, leaf blade, and sheath tissues as well as roots and germinating seeds (Figure 7). Transcripts included those of the 12 genes that lack *in vitro*

functional activity. There was little overlap in expression between above- and belowground tissues, indicating gene-specific adaptations in these tissues. Despite the observed transcriptional patterns, we were unable, with the exception of borneol, to detect volatile terpenes in leaves and roots of the Alamo cultivar under constitutive conditions.

To determine whether correlations between transcript abundance and volatile terpene products could be established in response to treatment with FAW, MeJA, and SA, we selected multiple TPS genes for expression analysis by quantitative RT-PCR (Figure 8A). In leaves, substantial induction at the transcript level (>10-fold) following herbivory was observed for 12 TPS genes (PvTPS01, PvTPS04, PvTPS05, PvTPS06, PvTPS08, PvTPS11, PvTPS14, PvTPS16, PvTPS19, PvTPS36, PvTPS53, and PvTPS56), of which 10 genes and PvTPS12 and PvTPS15 were equally of more highly induced upon root treatment with MeJA

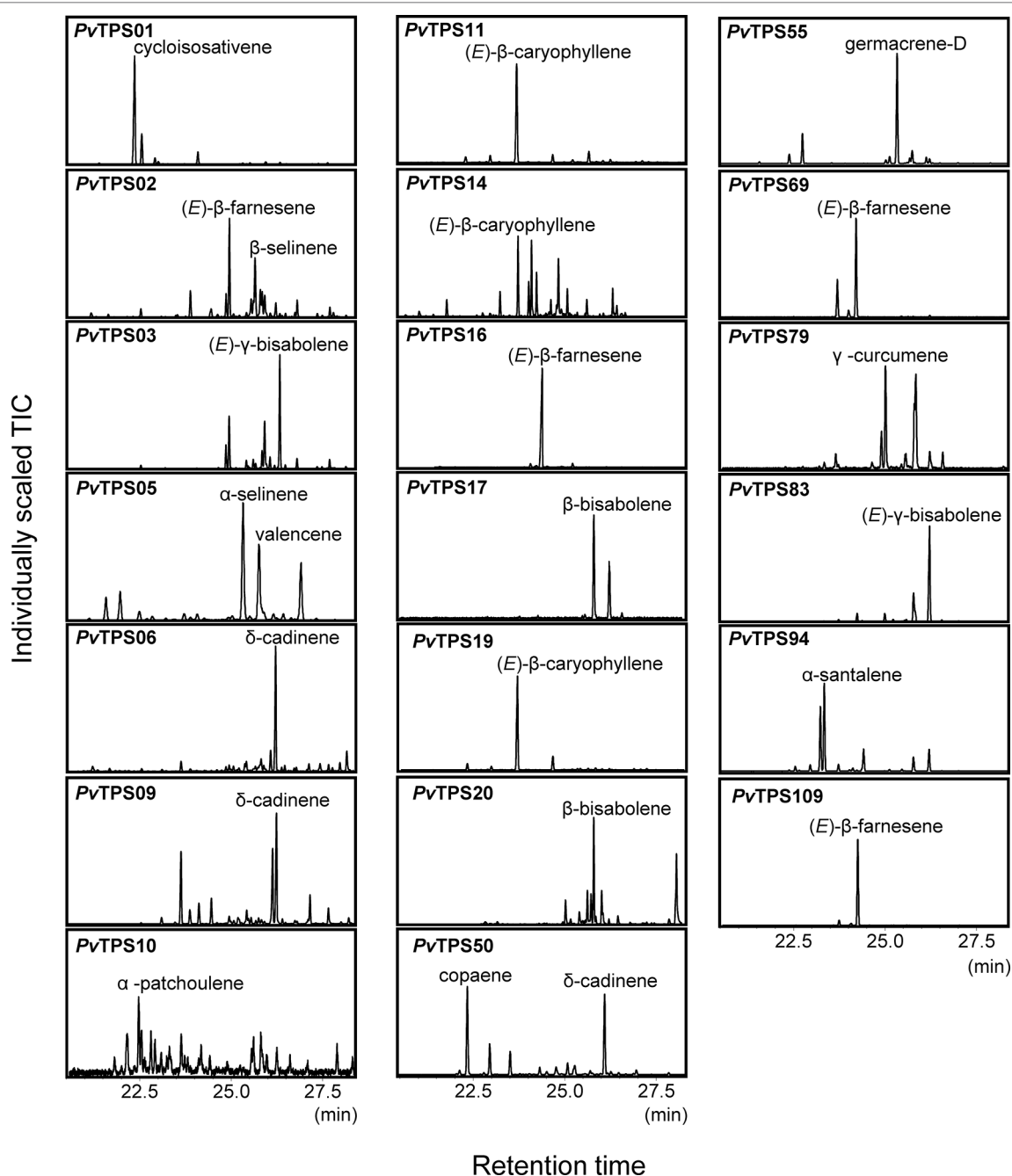


FIGURE 5 | Solid-phase microextraction–gas chromatography–mass spectrometry (SPME-GC-MS) analysis of sesquiterpenes produced by recombinant terpene synthases (TPSs) with farnesyl diphosphate [(*E,E*)-FDP] as a substrate. Compound identification was based on similarity to library matches (NIST, Wiley), authentic standards and comparison to Opopanax oil (*Commiphora guidotti*). Unlabeled peaks represent minor enzymatically produced putative sesquiterpene compounds.

(**Figure 8A**). SA-induced expression exceeding that in response to FAW and MeJA treatment was observed for *PvTPS04*, *PvTPS13*, *PvTPS16*, and *PvTPS53*. Highest induction of *TPS* transcript levels in roots was found for 11 genes in response to the application of MeJA (*PvTPS05*, *PvTPS06*, *PvTPS10*, *PvTPS11*, *PvTPS14*, *PvTPS17*, *PvTPS19*, *PvTPS20*, *PvTPS36*, *PvTPS53*, and *PvTPS56*) or both MeJA and SA (*PvTPS53*) (**Figure 8A**). For *PvTPS02*, *PvTPS03*, *PvTPS07*, and *PvTPS09*, induced transcript

levels were lower than 10-fold in leaves and/or roots upon any of the treatments (**Supplementary Figure S5**).

For nine *TPS* genes, we were able to identify their enzymatic products as components of the induced volatile blends of leaves and/or roots: The genes encoding (*E*)-β-caryophyllene synthases (*PvTPS11*, *PvTPS14*, and *PvTPS19*) showed highest transcript abundance in leaves and roots upon treatment with MeJA (**Figure 8A**). (*E*)-β-Caryophyllene emissions from both

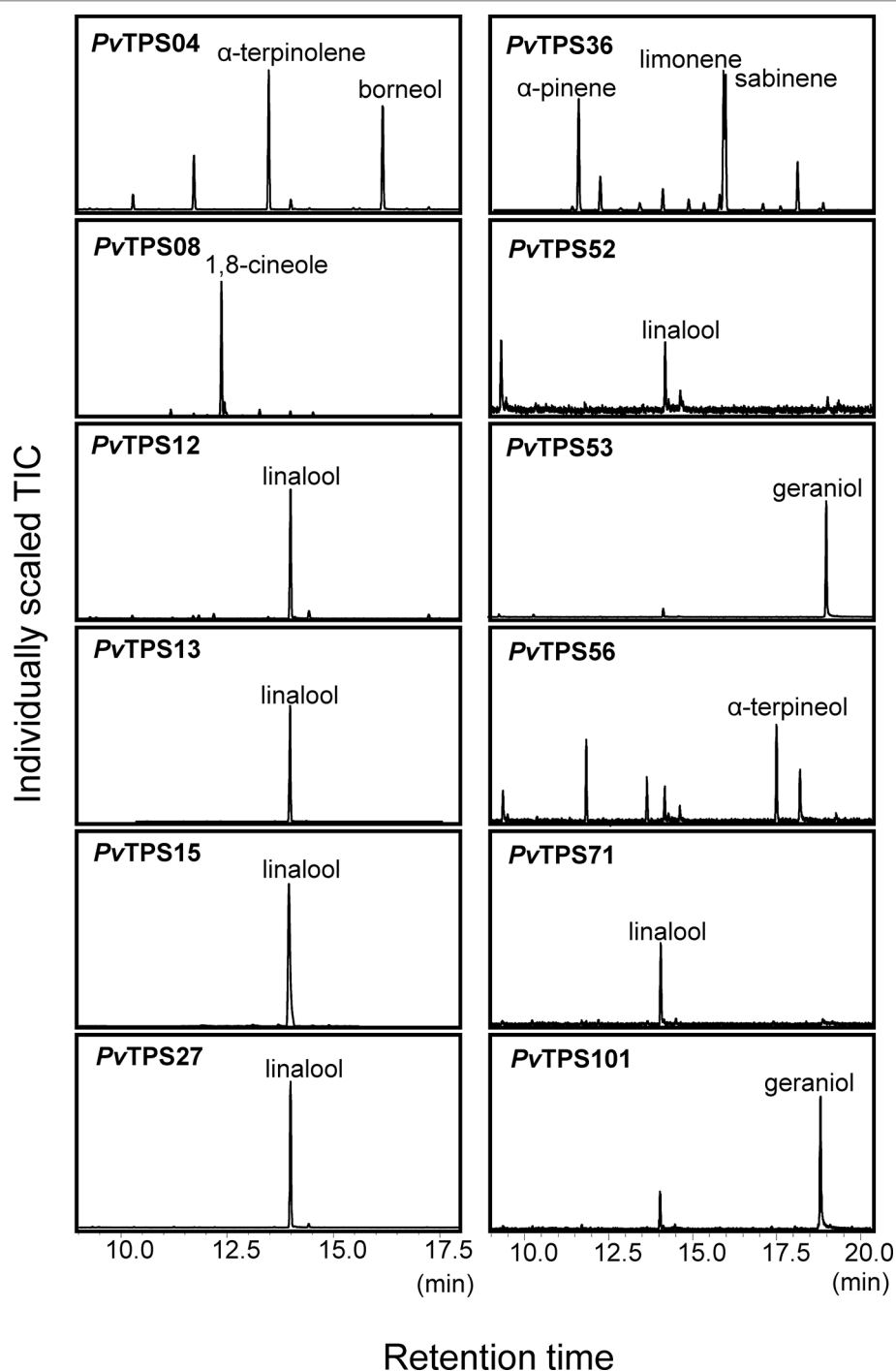


FIGURE 6 | Solid-phase microextraction–gas chromatography–mass spectrometry (SPME-GC-MS) analysis of volatile monoterpenes produced by truncated recombinant terpene synthases (TPSs) with geranyl diphosphate (GDP) as a substrate. Compound identification was based on similarity to library matches (NIST, Wiley), authentic standards, and comparison to Opopanax oil (*Commiphora guidotti*).

tissues are most likely associated with the activity of these TPSs. Expression of *PvTPS16*, whose recombinant protein produced (*E*)- β -farnesene, was strongly induced by SA treatment in leaves and is likely be associated with (*E*)- β -farnesene emission from this tissue (**Figures 1** and **8A**). The gene encoding *PvTPS05*,

which was found to produce α -selinene volatiles, was most strongly expressed in roots by application of MeJA matching the detection of this compound from root tissue (**Figures 2** and **8A**). *PvTPS06* and *PvTPS09* both catalyze the formation of δ -cadinene, which was emitted from roots upon MeJA treatment

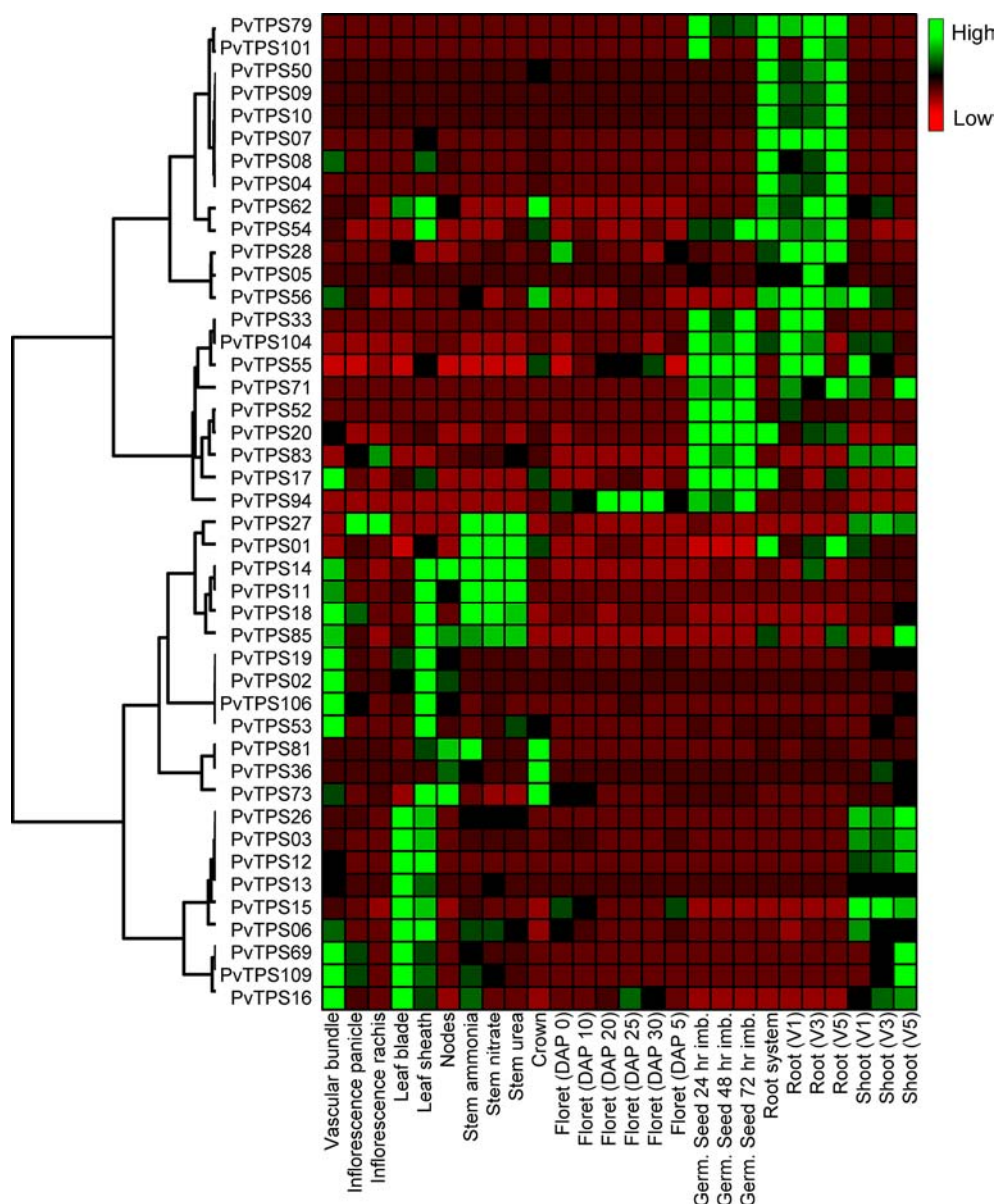


FIGURE 7 | Hierarchical cluster analysis of terpene synthase (TPS) gene expression in different switchgrass (cv. "AP13") tissues. The heat map compares relative transcript abundance for mono- and sesqui-TPS genes in fragments per kilobase of transcript per million mapped reads (FPKM) across 26 gene expression data sets (<https://phytozome.jgi.doe.gov/>).

(Figures 2 and 5). Only *PvTPS06* was highly induced by MeJA, indicating its likely function *in planta* (Figure 8A). Moreover, transcript levels of *PvTPS36* were substantially induced in leaves in response to MeJA application, although emission of one of the primary products of the *PvTPS36* enzyme, limonene, occurred only at low levels (Figures 1 and 8A). Interestingly, terpene products (cycloisoprenene, borneol) associated with two genes (*PvTPS01* and *PvTPS04*), which showed highest expression in leaves upon FAW and/or phytohormone treatment, could only be detected in roots (Figures 2 and 8A, Supplementary Figure S1).

Some TPS genes with lower levels of induction may contribute to the emission of particular terpenes [e.g., the (*E*)- β -farnesene

synthase gene *PvTPS02*]. Compounds produced *in vitro* by several other TPS enzymes could not be detected or occurred only at trace levels in leaves and roots despite a strong induction of their corresponding genes. For example, 1,8-cineole produced by *PvTPS08* was only detected in trace amounts in root tissues. Linalool, the single product of enzymes encoded by *PvTPS12*, *PvTPS13*, and *PvTPS15* (Figure 6), was neither detected in emissions from leaves and roots and may be further metabolized upon stress treatment. Other TPSs for which no associations could be established between their enzymatic products and volatile emissions include *PvTPS03* [(*E*)- γ -bisabolene synthase], *PvTPS10* (α -patchoulene synthase), *PvTPS17*, and *PvTPS20*

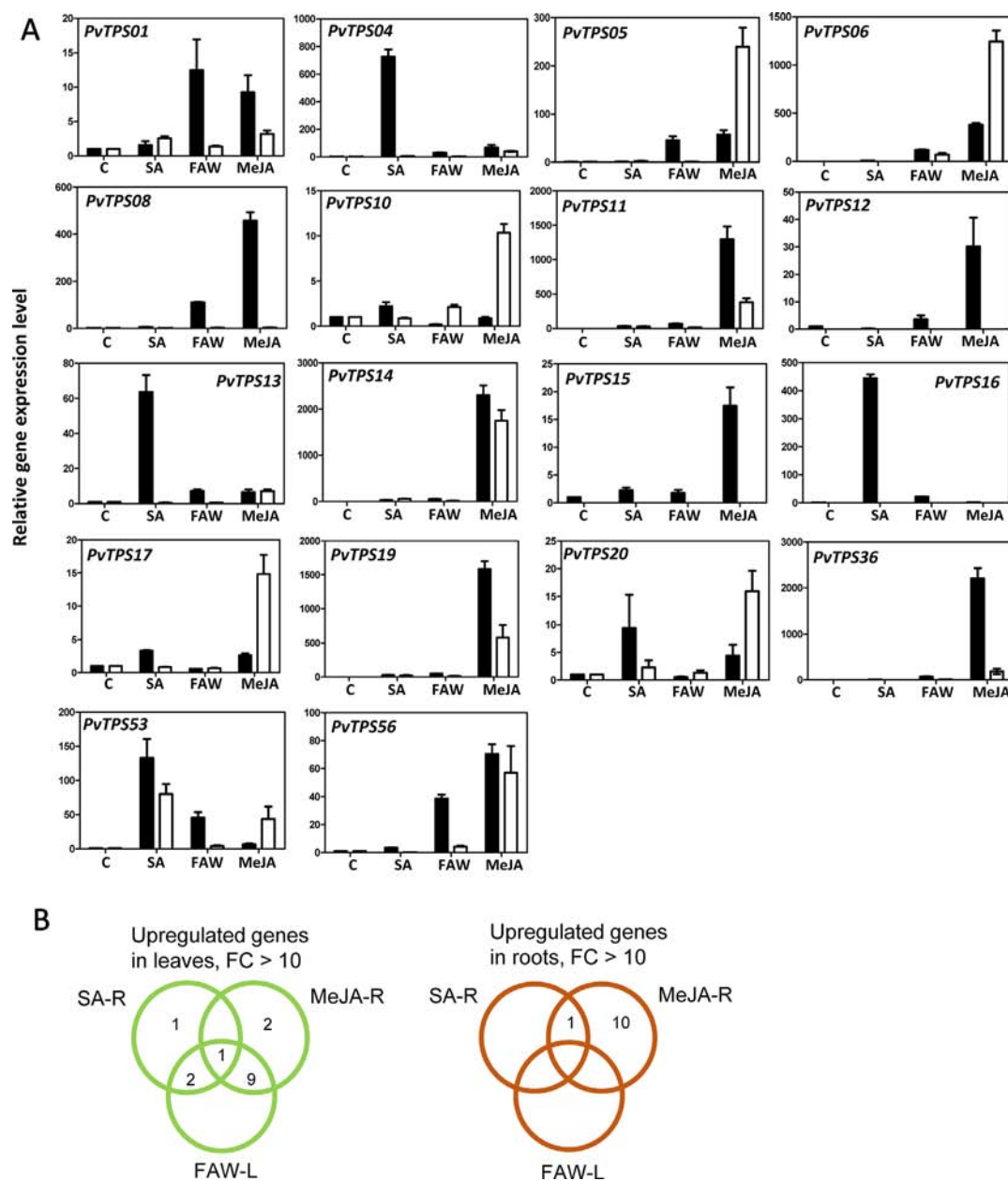


FIGURE 8 | Switchgrass terpene synthase (TPS) gene expression in response to different treatments. **(A)** Transcript levels of selected TPS genes induced >10-fold in leaves (black bars) and/or roots (white bars). Samples were analyzed in biological and technical triplicate and normalized to the expression of the control gene elongation factor 1a (*ELF1α*). Control (C) expression levels were scaled to 1 for comparison of treatment effects. **(B)** Venn diagrams of the TPS genes presented in **(A)** and induced by different treatments. R, root; L, leaf; SA, salicylic acid; FAW, fall armyworm; MeJA, methyl jasmonate.

[(*E*)- β -bisabolene synthases], and TPS53 (geraniol synthase). Emission of germacrene D from roots may be associated with *PvTPS55*, the expression of which was not determined.

DISCUSSION

The switchgrass genome contains a large family of 44 predicted full-length mono- and sesqui-TPS genes, of which 32 genes encode

functionally active proteins. Sesqui-TPSs belonging to the type-a subfamily make up the majority of this TPS group, while only few mono-TPSs have emerged in the type-a clade or are distributed over the type-b, type-g, and type-e/f clades (**Figure 3**). Similar distributions have been shown to occur in the TPS families of rice and sorghum (Chen et al., 2011). Expansions of type-a clades are also common in dicots, although typically a higher proportion of mono-TPSs can be found in the type-b and type-g clades of dicot species (Chen et al., 2011; Kulheim et al., 2015)

The size of the switchgrass mono- and sesqui-TPSs family is almost twice as large as the number of characterized or predicted proteins with mono- or sesqui-TPS activity in maize (Springer et al., 2018). Polyploidy likely contributed to the expansion of the switchgrass TPS gene family, which is in agreement with studies by Hofberger et al. (2015) demonstrating the role of polyploidy events in the diversification and expansion of terpene secondary metabolism. Gene duplication through polyploidization generates gene redundancy eventually increasing functional divergence and allowing species adaption (Wendel, 2000). As an allotetraploid, switchgrass evolved from two diploid ancestors giving rise to two complete subgenomes (N and K) and functional divergence of TPS genes. In *P. hallii*, a diploid relative of switchgrass, ~32 putative full-length TPS genes are annotated (<https://phytozome.jgi.doe.gov/>), indicating that polyploidization of switchgrass more than doubled the number of TPS genes. Polyploidy events in domesticated grasses may not always result in large TPS gene families as has been suggested for wheat (Schmelz et al., 2014). However, in switchgrass, obligate outcrossing and limited breeding have maintained massive phenotypic and adaptive polymorphisms (Casler et al., 2007), in line with a higher level of diversification in TPS genes. Nevertheless, one-third of the TPS genes we characterized appear to be functionally inactive, while several other TPSs might have limited *in vivo* activity due to their subcellular localization, suggesting inactivation and loss of *in vivo* function for a substantial fraction of the gene family.

A comparison between the switchgrass subgenomes found that only 35 or 50% of the TPS genes on subgenome K and N, respectively, have syntenic orthologs on the other subgenome. This limited synteny indicates subgenome divergence in TPS gene organization. Syntenic regions include TPS genes with identical functions [*PvTPS14* and *PvTPS19*—(*E*)- β -caryophyllene synthases; *PvTPS17* and *PvTPS20*—(*E*)- β -bisabolene synthases], while other orthologs adopted different functional activities. Further comparison with the genome of the closely related diploid species *P. hallii* revealed syntenic orthologs for more than 15 switchgrass TPS genes on 6 of the 9 *P. hallii* chromosomes. Corresponding syntenic orthologs could also be identified for several of these genes on the genomes of the close relative *S. italica* and of *S. bicolor*. These findings are consistent with the observed collinearity between the switchgrass, *Setaria*, and sorghum genomes (Casler et al., 2011) and suggest the presence of ancestral TPS genes in the common progenitor of sorghum and switchgrass more than 20 million years ago. Syntenic regions on the sorghum genome include a cluster of TPS genes on chromosome 7, which was found to encode insect-induced sesquiterpene synthases and shares (*E*)- β -farnesene synthase activity (*Sorbic.007G055600*, *PvTPS109*) (Zhuang et al., 2012).

Most mono- and sesqui-TPS genes of switchgrass exhibit tissue-specific expression patterns (Figure 7). With the exception of the root-accumulated monoterpene borneol, the products associated with these TPSs could not be found in leaves and roots under constitutive conditions and became in part only detectable in response to stress treatment when gene expression was induced. It is possible that under nontreatment conditions enzyme activity or substrate levels are too low to result in detectable amounts of

product. In roots, microbial activity may also metabolize terpene compounds as has been shown in vetiver grass (Del Giudice et al., 2008). It is also possible that the enzymatic products are further metabolized to nonvolatile derivatives. For example, β -macrocarpene, a volatile sesquiterpene olefin produced by two maize terpene synthases, is not detected in volatile blends because of its conversion to nonvolatile acid derivatives called zealexins, which function as pathogen-induced phytoalexins (Köllner et al., 2008b; Huffaker et al., 2011). In another study in maize, Ding et al. (2017) found that the volatile sesquiterpene β -selinene is a direct precursor of β -costic acid, a nonvolatile antibiotic acid derivative. Based on these findings, it is possible that α -selinene made by TPS05 in switchgrass roots serves as a precursor of α -costic acid that may exhibit similar functions in antimicrobial defense. Future analyses should be performed to identify possible oxygenated downstream derivatives of switchgrass TPS products.

Twelve TPS genes were found to be induced in switchgrass leaves upon feeding by FAW larvae. At least half of these genes are likely to contribute to the production of the volatile terpenes released upon FAW feeding based on the activity of their corresponding enzymes. The majority of the FAW-induced genes also responded to belowground treatment with MeJA, and two genes were induced by root treatment with SA indicating bottom-up systemic responses in *de novo* terpene biosynthesis (Figure 8B). While these effects are likely to be less pronounced with the application of lower concentrations of MeJA and SA or in response to actual root herbivory or pathogen infection, several studies have reported similar root-induced systemic responses in the metabolism of terpenoids and other secondary metabolites in photosynthetic tissues (Bezemer et al., 2003; Bezemer et al., 2004; Rasmann and Turlings, 2007; Erb et al., 2008; Kaplan et al., 2008). By contrast, much weaker systemic effects have been observed on root defensive metabolites including terpenes in maize upon shoot treatments or foliar feeding (Bezemer et al., 2003; Bezemer et al., 2004; Rasmann and Turlings, 2007; Erb et al., 2008; Kaplan et al., 2008). Our findings support this notion since FAW feeding did not cause a major increase in TPS gene expression in switchgrass roots and only a local treatment with MeJA could elicit such a response (Figure 8B).

The terpene olefins released by switchgrass leaves and roots upon insect or hormone treatment are frequently found in stress-induced volatile blends of other monocots and dicots (Unsicker et al., 2009; Massalha et al., 2017). While determining the function of these compounds is beyond the scope of this study, we assume that they play roles in direct and indirect defenses similar to those described previously in maize, rice, or other plants (Degenhardt et al., 2009; Hare and Sun, 2011; Taniguchi et al., 2014; Chen et al., 2018). A common constituent of herbivore-induced volatile blends in many plants including grasses is (*E*)- β -caryophyllene (Köllner et al., 2008a). This sesquiterpene, when released from damaged leaves of maize and rice plants, has been implicated in recruiting parasitoids of herbivores (Cheng et al., 2007; Köllner et al., 2008a; Yuan et al., 2008). We identified three (*E*)- β -caryophyllene synthase genes (*PvTPS11*, *PvTPS14*, and *PvTPS19*) (Figure 5), all of which are located on chromosome 2 and induced upon FAW feeding and treatment with MeJA. By contrast, in maize, rice, and sorghum, only single genes

(*ZmTPS23*, *Os08g04500*, *SbTPS4*) have been associated with the synthesis of (*E*)- β -caryophyllene upon herbivore feeding (Köllner et al., 2008a; Zhuang et al., 2012; Chen et al., 2014). In MeJA-treated root tissue, *PvTPS14* was found to be induced approximately fourfold higher than *PvTPS11* and *PvTPS19* and may contribute to the emission of (*E*)- β -caryophyllene belowground. Induced root expression of (*E*)- β -caryophyllene synthases is common among grasses and has been implicated with recruitment of entomopathogenic nematodes for indirect defense against belowground herbivory (Rasmann et al., 2005).

(*E*)- β -Farnesene is another sesquiterpene that is released by many plant species and plays, among other volatiles, a role in indirect defense in maize (Schnee et al., 2006; Degenhardt, 2009). We found four TPS genes that encode functionally active (*E*)- β -farnesene synthases (Figure 5). However, only *PvTPS02* expression correlated with compound emission as a result of herbivore damage (Figure 1 and Supplementary Figure S5). Another gene, *PvTPS16*, was highly expressed in leaves following SA treatment and strongly correlated with (*E*)- β -farnesene emission under this condition (Figures 1 and 8A). Despite limited and controversial evidence (Gibson and Pickett, 1983; Kunert et al., 2010), this response could potentially affect aphids, since (*E*)- β -farnesene serves as an alarm pheromone for many aphid taxa (Bowers et al., 1977; Pickett, 1983) and aphids are known to elicit both SA- and JA-dependent signaling pathways (Moran et al., 2002). A recent study by Donze-Reiner et al. (2017) found several TPS genes to be induced upon feeding by the grain aphid *S. graminum*. However, none of the (*E*)- β -farnesene synthase genes was among those induced by *S. graminum*, indicating that their expression might be suppressed. Instead, genes induced by aphid feeding included the (*E*)- β -bisabolene synthases *PvTPS17* and *PvTPS20* among other genes in the type-a family and genes in the type-c and type-e/f families, which have in part be characterized as diterpene synthases (Pelot et al., 2018). Whether these terpene compounds are produced upon *S. graminum* feeding is currently unknown.

We found only two monoterpenes (limonene and β -ocimene) to be emitted at low levels from treated switchgrass leaves (Figure 1). Except of *PvTPS36*, which was induced in leaves by MeJA treatment and makes limonene as an enzymatic product (Figures 6 and 8A), no terpene products of the other induced mono-TPS genes could be detected possibly because of the reasons addressed earlier. Interestingly, enzymatic products of two TPSs, the cycloisotriterpene synthase *PvTPS01* and the borneol synthase, *PvTPS04*, could only be observed in emissions from roots, although the corresponding genes were most highly expressed in leaves upon FAW, MeJA, or SA treatment (Figures 1 and 8A, Supplementary Figure S1). Whether the absence of the compounds in leaf tissue is due to limited enzymatic activity, metabolism of the product, or transport from shoots to roots remains to be determined.

In summary, our study has provided a genetic road map for investigating the biosynthesis and function of volatile terpenoids in switchgrass. We have shown that the switchgrass genome contains an extended family of mono- and sesqui-TPS genes, several of which share syntenic orthologs in other grasses,

exhibit tissue-specific expression, and respond to herbivory and phytohormone treatment above- and belowground. The volatiles associated with these genes and possibly their nonvolatile derivatives may exhibit functions in above- and belowground direct and indirect defense similar to those described for maize and other grasses. Further studies involving the generation of switchgrass mutants will evaluate these ecological roles in greater detail.

DATA AVAILABILITY

The datasets generated for this study can be found in Phytozome, <https://phytozome.jgi.doe.gov/pz/portal.html>.

AUTHOR CONTRIBUTIONS

AM, XC, FC, and DT designed the study. AM, XC, TK, KP, and PZ performed bioinformatic analyses and gene annotation. JL performed synteny analyses. AM, XC, MR, LC, and SL performed enzyme characterizations. AM and XC performed RNA extraction, RT-qPCR, and stress treatments. AM and XC performed volatile profiling. AM, XC, FC, and DT wrote the manuscript. All authors reviewed, read, and approved the manuscript before submission.

FUNDING

This work was supported by Community Science Program grant (WIP 2568) of the Department of Energy Joint Genome Institute and funds by the Translational Plant Sciences Program at Virginia Tech. The work conducted by the US Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the US Department of Energy under contract no. DE-AC02-05CH11231.

ACKNOWLEDGMENTS

We would like to thank Dr. Nan Zhao for his assistance with volatile profiling and Dr. Qidong Jia for his assistant with sequence analysis. We thank Debbie Wiley for assistance with plant maintenance. We are grateful to Dr. Gerald A. Tuskan (Center for Bioenergy Innovation, Oakridge National Laboratory) and Dr. Yuhong Tang of the Noble Foundation for the availability and access to switchgrass RNA-seq data sets. We thank Jim Tokuhisa for scientific advice. We thank the US Department of Energy Joint Genome Institute and collaborators for prepublication access to the *Panicum virgatum* V1.1 and V4.1 genome sequence.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.01144/full#supplementary-material>

REFERENCES

- Bezemer, T. M., Wagenaar, R., Van Dam, N. M., and Wackers, F. L. (2003). Interactions between above- and belowground insect herbivores as mediated by the plant defense system. *Oikos* 101, 555–562. doi: 10.1034/j.1600-0706.2003.12424.x
- Bezemer, T. M., Wagenaar, R., Van Dam, N. M., Van Der Putten, W. H., and Wackers, F. L. (2004). Above- and below-ground terpenoid aldehyde induction in cotton, *Gossypium herbaceum*, following root and leaf injury. *J. Chem. Ecol.* 30, 53–67. doi: 10.1023/B:JOEC.0000013182.50662.2a
- Block, A. K., Vaughan, M. M., Schmelz, E. A., and Christensen, S. A. (2019). Biosynthesis and function of terpenoid defense compounds in maize (*Zea mays*). *Planta* 249, 21–30. doi: 10.1007/s00425-018-2999-2
- Boutanaev, A. M., Moses, T., Zi, J. C., Nelson, D. R., Mugford, S. T., Peters, R. J., et al. (2015). Investigation of terpene diversification across multiple sequenced plant genomes. *Proc. Natl. Acad. Sci. U.S.A.* 112, E81–E88. doi: 10.1073/pnas.1419547112
- Bowers, W. S., Nishino, C., Montgomery, M. E., and Nault, L. R. (1977). Structure-activity relationships of analogs of the aphid alarm pheromone, (*E*)- β -farnesene. *J. Insect Physiol.* 23, 697–701. doi: 10.1016/0022-1910(77)90086-5
- Casler, M. D., Tobias, C. M., Kaeppeler, S. M., Buell, C. R., Wang, Z. Y., Cao, P. J., et al. (2011). The switchgrass genome: tools and strategies. *Plant Genome* 4, 273–282. doi: 10.3835/plantgenome2011.10.0026
- Casler, M. D., Vogel, K. P., Taliaferro, C. M., Ehler, N. J., Berdahl, J. D., Brummer, E. C., et al. (2007). Latitudinal and longitudinal adaptation of switchgrass populations. *Crop Sci.* 47, 2249–2260. doi: 10.2135/cropsci2006.12.0780
- Chen, X. J., Chen, H., Yuan, J. S., Köllner, T. G., Chen, Y. Y., Guo, Y. F., et al. (2018). The rice terpene synthase gene OsTPS19 functions as an (*S*)-limonene synthase in planta, and its overexpression leads to enhanced resistance to the blast fungus *Magnaporthe oryzae*. *Plant Biotechnol. J.* 16, 1778–1787. doi: 10.1111/pbi.12914
- Chen, H., Li, G. L., Köllner, T. G., Jia, Q. D., Gershenzon, J., and Chen, F. (2014). Positive Darwinian selection is a driving force for the diversification of terpenoid biosynthesis in the genus *Oryza*. *BMC Plant Biol.* 14, 239. doi: 10.1186/s12870-014-0239-x
- Chen, F., Tholl, D., Bohlmann, J., and Pichersky, E. (2011). The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *Plant J.* 66, 212–229. doi: 10.1111/j.1365-3113.2011.04520.x
- Cheng, A. X., Xiang, C. Y., Li, J. X., Yang, C. Q., Hu, W. L., Wang, L. J., et al. (2007). The rice (*E*)- β -caryophyllene synthase (OsTPS3) accounts for the major inducible volatile sesquiterpenes. *Phytochemistry* 68, 1632–1641. doi: 10.1016/j.phytochem.2007.04.008
- Degenhardt, J. (2009). Indirect defense responses to herbivory in grasses. *Plant Physiol.* 149, 96–102. doi: 10.1104/pp.108.128975
- Degenhardt, J., Hiltbold, I., Köllner, T. G., Frey, M., Gierl, A., Gershenzon, J., et al. (2009). Restoring a maize root signal that attracts insect-killing nematodes to control a major pest. *Proc. Natl. Acad. Sci. U.S.A.* 106, 13213–13218. doi: 10.1073/pnas.0906365106
- Del Giudice, L., Massardo, D. R., Pontieri, P., Berte, C. M., Mombello, D., Carata, E., et al. (2008). The microbial community of Vetiver root and its involvement into essential oil biogenesis. *Environ. Microbiol.* 10, 2824–2841. doi: 10.1111/j.1462-2920.2008.01703.x
- Dicke, M. (1994). Local and systemic production of volatile herbivore-induced terpenoids—their role in plant–carnivore mutualism. *J. Plant Physiol.* 143, 465–472. doi: 10.1016/S0176-1617(11)81808-0
- Ding, Y. Z., Huffaker, A., Köllner, T. G., Weckwerth, P., Robert, C. A. M., Spencer, J. L., et al. (2017). Selenine volatiles are essential precursors for maize defense promoting fungal pathogen resistance. *Plant Physiol.* 175, 1455–1468. doi: 10.1104/pp.17.00879
- Donze-Reiner, T., Palmer, N. A., Scully, E. D., Prochaska, T. J., Koch, K. G., Heng-Moss, T., et al. (2017). Transcriptional analysis of defense mechanisms in upland tetraploid switchgrass to greenbugs. *BMC Plant Biol.* 17, 46. doi: 10.1186/s12870-017-0998-2
- Dudareva, N., Pichersky, E., and Gershenzon, J. (2004). Biochemistry of plant volatiles. *Plant Physiol.* 135, 1893–1902. doi: 10.1104/pp.104.049981
- Emanuelsson, O., Nielsen, H., and Von Heijne, G. (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci.* 8, 978–984. doi: 10.1110/ps.8.5.978
- Erb, M., Ton, J., Degenhardt, J., and Turlings, T. C. J. (2008). Interactions between arthropod-induced aboveground and belowground defenses in plants. *Plant Physiol.* 146, 867–874. doi: 10.1104/pp.107.112169
- Erb, M., Veyrat, N., Robert, C. A. M., Xu, H., Frey, M., Ton, J., et al. (2015). Indole is an essential herbivore-induced volatile priming signal in maize. *Nat. Commun.* 6, 6273. doi: 10.1038/ncomms7273
- Gibson, R. W., and Pickett, J. A. (1983). Wild potato repels aphids by release of aphid alarm pheromone. *Nature* 302, 608–609. doi: 10.1038/302608a0
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–U130. doi: 10.1038/nbt.1883
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010
- Hare, J. D., and Sun, J. J. (2011). Production of induced volatiles by *Datura wrightii* in response to damage by insects: effect of herbivore species and time. *J. Chem. Ecol.* 37, 751–764. doi: 10.1007/s10886-011-9985-5
- Hofberger, J. A., Ramirez, A. M., van den Bergh, E., Zhu, X. G., Bouwmeester, H. J., Schuurink, R. C., et al. (2015). Large-scale evolutionary analysis of genes and supergene clusters from terpenoid modular pathways provides insights into metabolic diversification in flowering plants. *PLoS One* 10, e0128808. doi: 10.1371/journal.pone.0128808
- Huang, M., Sanchez-Moreiras, A. M., Abel, C., Sohrabi, R., Lee, S., Gershenzon, J., et al. (2012). The major volatile organic compound emitted from *Arabidopsis thaliana* flowers, the sesquiterpene (*E*)- β -caryophyllene, is a defense against a bacterial pathogen. *New Phytol.* 193, 997–1008. doi: 10.1111/j.1469-8113.2011.04001.x
- Huffaker, A., Kaplan, F., Vaughan, M. M., Dafoe, N. J., Ni, X. Z., Rocca, J. R., et al. (2011). Novel acidic sesquiterpenoids constitute a dominant class of pathogen-induced phytoalexins in maize. *Plant Physiol.* 156, 2082–2097. doi: 10.1104/pp.111.179457
- Kaplan, I., Halitschke, R., Kessler, A., Sardaneli, S., and Denno, R. F. (2008). Constitutive and induced defenses to herbivory in above- and belowground plant tissues. *Ecology* 89, 392–406. doi: 10.1890/07-0471.1
- Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436
- Köllner, T. G., Held, M., Lenk, C., Hiltbold, I., Turlings, T. C. J., Gershenzon, J., et al. (2008a). A maize (*E*)- β -caryophyllene synthase implicated in indirect defense responses against herbivores is not expressed in most American maize varieties. *Plant Cell* 20, 482–494. doi: 10.1105/tpc.107.051672
- Köllner, T. G., Schnee, C., Li, S., Svatos, A., Schneider, B., Gershenzon, J., et al. (2008b). Protonation of a neutral (*S*)- β -bisabolene intermediate is involved in (*S*)- β -macrocarypene formation by the maize sesquiterpene synthases TPS6 and TPS11. *J. Biol. Chem.* 283, 20779–20788. doi: 10.1074/jbc.M802682200
- Kost, C., and Heil, M. (2006). Herbivore-induced plant volatiles induce an indirect defence in neighbouring plants. *J. Ecol.* 94, 619–628. doi: 10.1111/j.1365-2745.2006.01120.x
- Kulheim, C., Padovan, A., Hefer, C., Krause, S. T., Köllner, T. G., Myburg, A. A., et al. (2015). The *Eucalyptus* terpene synthase gene family. *BMC Genomics* 16, 450. doi: 10.1186/s12864-015-1598-x
- Kunert, G., Reinhold, C., and Gershenzon, J. (2010). Constitutive emission of the aphid alarm pheromone, (*E*)- β -farnesene, from plants does not serve as a direct defense against aphids. *BMC Ecol.* 10, 23. doi: 10.1186/1472-6785-10-23
- Lee, S. T., Mitchell, R. B., Wang, Z., Heiss, C., Gardner, D. R., and Azadi, P. (2009). Isolation, characterization, and quantification of steroidal saponins in switchgrass (*Panicum virgatum* L.). *J. Agric. Food Chem.* 57, 2599–2604. doi: 10.1021/jf803907y
- Letunic, I., and Bork, P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23, 127–128. doi: 10.1093/bioinformatics/btl529
- Lovell, J. T., Jenkins, J., Lowry, D. B., Mamidi, S., Sreedasyam, A., Weng, X. Y., et al. (2018). The genomic landscape of molecular responses to natural drought stress in *Panicum hallii*. *Nat. Commun.* 9, 5213. doi: 10.1038/s41467-018-07669-x

- Massalha, H., Korenblum, E., Tholl, D., and Aharoni, A. (2017). Small molecules below-ground: the role of specialized metabolites in the rhizosphere. *Plant J.* 90, 788–807. doi: 10.1111/tjp.13543
- Moran, P. J., Cheng, Y., Cassell, J. L., and Thompson, G. A. (2002). Gene expression profiling of *Arabidopsis thaliana* in compatible plant–aphid interactions. *Arch. Insect Biochem. Physiol.* 51, 182–203. doi: 10.1002/arch.10064
- Osbourne, A. E. (1996). Preformed antimicrobial compounds and plant defense against fungal attack. *Plant Cell* 8, 1821–1831. doi: 10.1105/tpc.8.10.1821
- Parrish, D. J., and Fike, J. H. (2005). The biology and agronomy of switchgrass for biofuels. *Crit. Rev. Plant Sci.* 24, 423–459. doi: 10.1080/07352680500316433
- Pelot, K. A., Chen, R., Hagelthorn, D. M., Young, C. A., Addison, J. B., Muchlinski, A., et al. (2018). Functional diversity of diterpene synthases in the biofuel crop switchgrass. *Plant Physiol.* 178, 54–71. doi: 10.1104/pp.18.00590
- Pichersky, E., and Gang, D. R. (2000). Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. *Trends Plant Sci.* 5, 439–445. doi: 10.1016/S1360-1385(00)01741-6
- Pickett, R. W. G. J. A. (1983). Wild potato repels aphids by release of aphid alarm pheromone. *Nature*, 302, 608–609. doi: 10.1038/302608a0
- Rasmann, S., and Turlings, T. C. J. (2007). Simultaneous feeding by aboveground and belowground herbivores attenuates plant-mediated attraction of their respective natural enemies. *Ecol. Lett.* 10, 926–936. doi: 10.1111/j.1461-0248.2007.01084.x
- Rasmann, S., Köllner, T. G., Degenhardt, J., Hiltbold, I., Toepfer, S., Kuhlmann, U., et al. (2005). Recruitment of entomopathogenic nematodes by insect-damaged maize roots. *Nature* 434, 732–737. doi: 10.1038/nature03451
- Sanderson, M. A., Adler, P. R., Boateng, A. A., Casler, M. D., and Sarath, G. (2006). Switchgrass as a biofuels feedstock in the USA. *Can. J. Plant Sci.* 86, 1315–1325. doi: 10.4141/P06-136
- Schmelz, E. A., Huffaker, A., Sims, J. W., Christensen, S. A., Lu, X., Okada, K., et al. (2014). Biosynthesis, elicitation and roles of monocot terpenoid phytoalexins. *Plant J.* 79, 659–678. doi: 10.1111/tjp.12436
- Schnee, C., Köllner, T. G., Held, M., Turlings, T. C. J., Gershenzon, J., and Degenhardt, J. (2006). The products of a single maize sesquiterpene synthase form a volatile defense signal that attracts natural enemies of maize herbivores. *Proc. Natl. Acad. Sci. U.S.A.* 103, 1129–1134. doi: 10.1073/pnas.0508027103
- Springer, N. M., Anderson, S. N., Andorf, C. M., Ahern, K. R., Bai, F., Barad, O., et al. (2018). The maize W22 genome provides a foundation for functional genomics and transposon biology. *Nat. Genet.* 50, 1282–1288. doi: 10.1038/s41588-018-0158-0
- Taniguchi, S., Hosokawa-Shinonaga, Y., Tamaoki, D., Yamada, S., Akimitsu, K., and Gomi, K. (2014). Jasmonate induction of the monoterpene linalool confers resistance to rice bacterial blight and its biosynthesis is regulated by JAZ protein in rice. *Plant Cell Environ.* 37, 451–461. doi: 10.1111/pce.12169
- Tholl, D., and Lee, S. (2011). Terpene specialized metabolism in *Arabidopsis thaliana*. *Arabidopsis Book* 9, e0143. doi: 10.1199/tab.0143
- Tholl, D., Chen, F., Petri, J., Gershenzon, J., and Pichersky, E. (2005). Two sesquiterpene synthases are responsible for the complex mixture of sesquiterpenes emitted from *Arabidopsis* flowers. *Plant J.* 42, 757–771. doi: 10.1111/j.1365-313X.2005.02417.x
- Turlings, T. C. J., Tumlinson, J. H., and Lewis, W. J. (1990). Exploitation of herbivore-induced plant odors by host-seeking parasitic wasps. *Science* 250, 1251–1253. doi: 10.1126/science.250.4985.1251
- Unsicker, S. B., Kunert, G., and Gershenzon, J. (2009). Protective perfumes: the role of vegetative volatiles in plant defense against herbivores. *Curr. Opin. Plant Biol.* 12, 479–485. doi: 10.1016/j.pbi.2009.04.001
- Vaughan, M. M., Wang, Q., Webster, F. X., Kiemle, D., Hong, Y. J., Tantillo, D. J., et al. (2013). Formation of the unusual semivolatile diterpene rhizathalene by the *Arabidopsis* class I terpene synthase TPS08 in the root stele is involved in defense against belowground herbivory. *Plant Cell* 25, 1108–1125. doi: 10.1105/tpc.112.100057
- Vogel, K. P. (2004). “Switchgrass,” in *Warm-season (C4) grasses*. Eds. L. E. Moser, B. L. Burson and L. E. Sollenberger (Madison, WI: American Society of Agronomy, Crop Science Society of America, Soil Science Society of America).
- Wendel, J. F. (2000). Genome evolution in polyploids. *Plant Mol. Biol.* 42, 225–249. doi: 10.1023/A:1006392424384
- Yuan, J. S., Köllner, T. G., Wiggins, G., Grant, J., Degenhardt, J., and Chen, F. (2008). Molecular and genomic basis of volatile-mediated indirect defense against insects in rice. *Plant J.* 55, 491–503. doi: 10.1111/j.1365-313X.2008.03524.x
- Zhuang, X. F., Köllner, T. G., Zhao, N., Li, G. L., Jiang, Y. F., Zhu, L. C., et al. (2012). Dynamic evolution of herbivore-induced sesquiterpene biosynthesis in sorghum and related grass crops. *Plant J.* 69, 70–80. doi: 10.1111/j.1365-313X.2011.04771.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Muchlinski, Chen, Lovell, Köllner, Pelot, Zerbe, Ruggiero, Callaway, Laliberte, Chen and Tholl. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Spatio-Temporal Metabolite and Elemental Profiling of Salt Stressed Barley Seeds During Initial Stages of Germination by MALDI-MSI and μ -XRF Spectrometry

Sneha Gupta¹, Thusitha Rupasinghe², Damien L. Callahan³, Siria H. A. Natera², Penelope M. C. Smith⁴, Camilla B. Hill⁵, Ute Roessner² and Berin A. Boughton^{1,2*}

OPEN ACCESS

Edited by:

Kazuki Saito,
RIKEN Center for Sustainable
Resource Science (CSRS),
Japan

Reviewed by:

Robert David Hall,
Wageningen University & Research,
Netherlands
Yozo Okazaki,
Mie University,
Japan

*Correspondence:

Berin A. Boughton
baboug@unimelb.edu.au

Specialty section:

This article was submitted to
Plant Metabolism and
Chemodiversity,
a section of the journal
Frontiers in Plant Science

Received: 05 April 2019

Accepted: 21 August 2019

Published: 25 September 2019

Citation:

Gupta S, Rupasinghe T, Callahan DL,
Natera SHA, Smith PMC, Hill CB,
Roessner U and Boughton BA (2019)
Spatio-Temporal Metabolite and
Elemental Profiling of Salt Stressed
Barley Seeds During Initial Stages
of Germination by MALDI-MSI and
 μ -XRF Spectrometry.
Front. Plant Sci. 10:1139.
doi: 10.3389/fpls.2019.01139

¹ School of BioSciences, University of Melbourne, Parkville, VIC, Australia, ² Metabolomics Australia, School of BioSciences, University of Melbourne, Parkville, VIC, Australia, ³ School of Life and Environmental Sciences, Deakin University, Burwood, VIC, Australia, ⁴ AgriBio, Centre for AgriBiosciences, Department of Animal, Plant and Soil Sciences, School of Life Sciences, La Trobe University, Bundoora, VIC, Australia, ⁵ School of Veterinary and Life Sciences, Murdoch University, Murdoch, WA, Australia

Seed germination is the essential first step in crop establishment, and can be severely affected by salinity stress which can inhibit essential metabolic processes during the germination process. Salt stress during seed germination can trigger lipid-dependent signalling cascades that activate plant adaptation processes, lead to changes in membrane fluidity to help resist the stress, and cause secondary metabolite responses due to increased oxidative stress. In germinating barley (*Hordeum vulgare*), knowledge of the changes in spatial distribution of lipids and other small molecules at a cellular level in response to salt stress is limited. In this study, mass spectrometry imaging (MSI), liquid chromatography quadrupole time-of-flight mass spectrometry (LC-QToF-MS), inductively coupled plasma mass spectrometry (ICP-MS), and X-ray fluorescence (XRF) were used to determine the spatial distribution of metabolites, lipids and a range of elements, such as K⁺ and Na⁺, in seeds of two barley genotypes with contrasting germination phenology (Australian barley varieties Mundah and Keel). We detected and tentatively identified more than 200 lipid species belonging to seven major lipid classes (fatty acyls, glycerolipids, glycerophospholipids, sphingolipids, prenol lipids, sterol lipids, and polyketides) that differed in their spatial distribution based on genotype (Mundah or Keel), time post-imbibition (0 to 72 h), or treatment (control or salt). We found a tentative flavonoid was discriminant in post-imbibed Mundah embryos under saline conditions, and a delayed flavonoid response in Keel relative to Mundah. We further employed MSI-MS/MS and LC-QToF-MS/MS to explore the identity of the discriminant flavonoid and study the temporal pattern in five additional barley genotypes. ICP-MS was used to quantify the elemental composition of both Mundah and Keel seeds, showing a significant increase in Na⁺ in salt treated samples. Spatial mapping of elements using μ -XRF localized the elements within the seeds. This study integrates data obtained from three mass

spectrometry platforms together with μ -XRF to yield information on the localization of lipids, metabolites and elements improving our understanding of the germination process under salt stress at a molecular level.

Keywords: salinity, barley, germination, metabolomics, mass spectrometry imaging, lipids, MALDI

INTRODUCTION

Barley is a model organism for investigating the cereal germination process (Gorzolka et al., 2016) which commences with the uptake of water by the quiescent dry seed and finishes with the emergence of the radicle through the seed coat. Seed germination requires sufficient moisture and is also affected by temperature. Salinity affects normal seed germination through osmotic stress (Bliss et al., 1986), ion toxicity (Hampson and Simpson, 1990), or a combination of both (Huang and Steveninck, 1988). High intracellular concentrations of both Na^+ and Cl^- can inhibit the metabolism of dividing and expanding cells (Neumann, 1997), restrict mobilization, and hinder seedling emergence (Marques et al., 2013; Alencar et al., 2015). These conditions result in retarded (Zhang et al., 2010) or delayed (Ashraf et al., 2003) germination. Barley genotypes can be classified as either tolerant or sensitive to salt stress depending upon their genetic diversity and the ability to germinate and, survive under these conditions (Shelden and Roessner, 2013; Shelden et al., 2013; Shelden et al., 2016). The molecular changes in barley seed during germination that may play a role in tolerance or sensitivity are important in determining a plant's overall response to salinity.

Mature seeds contain a variety of compounds such as proteins, carbohydrates, lipids, vitamins, and phenolics. These compounds provide essential nutrients for the seed to germinate and develop into a mature plant, and are synthesized, packaged and stored in specific tissues that are chemically and morphologically distinct from each other (Fulcher, 1982). For example, a reserve of carbohydrates (starch) is found in the endosperm of cereal seed whereas a reserve of lipids and proteins is found in the scutellum.

Lipids are vital and abundant cellular constituents responsible for the structure and function of cell membranes, and act as an energy store to allow metabolism to continue during abiotic stress. Recent research has shown that lipids are involved as signal mediators in the initiation of defence reactions (Sarabia et al., 2018). Lipids are also involved in processes to mitigate the effects of stress in plant cells (Hasanuzzaman et al., 2013; Okazaki and Saito, 2014) for example through remodelling of glycerolipid levels to maintain membrane integrity and optimal fluidity (Sarabia et al., 2018). Plants also produce a vast array of secondary metabolites such as carotenoids, flavonoids, and coumestans (Gry et al., 2007). These compounds have long been known to protect plants from different biotic and abiotic stresses, function as signal molecules and act as antimicrobial agents (Bartwal et al., 2013).

Although changes in profiles of protein (Yang et al., 2007), transcripts (Nakabayashi et al., 2005; Reyes and Chua, 2007), and metabolites (Fait et al., 2006; Howell et al., 2009) during germination are well documented, there is currently little

information on molecular changes during seed germination in response to abiotic stress including salinity. With the current changes in the global climate (Karl et al., 2009), understanding these responses to stress will be more important than ever. Information on how an embryo mobilizes its internal reserves during the early stages of germination in saline conditions can provide insights into the metabolic process of germination; and consequently help us understand how certain crop genotypes germinate better under adverse environmental conditions (Achakzai, 2014).

To understand the role and function of metabolites and lipids within an organism at the tissue level, spatial information about target compounds is required (Li et al., 2008). Many studies analyze whole tissue responses (Fiehn, 2002; Sumner et al., 2003; Weckwerth, 2003; Cevallos-Cevallos et al., 2009; Lee et al., 2010), and so information about the localization of small molecules in cell types within the tissue is lost.

Mass Spectrometry Imaging (MSI) provides a platform to detect and localize intra-tissue variation and the spatial distribution of metabolites at the cellular and sub-cellular levels at high resolution (Lee et al., 2012; Boughton et al., 2016). MSI enables visualization of the distribution of individual biomolecules in a tissue section without requiring staining or complicated pretreatment (Cornett et al., 2007). This method of direct analysis of a tissue section (Zaima et al., 2009) allows detection of a wide range of endogenous molecules such as lipids, peptides, and secondary metabolites (Boughton et al., 2016; Hansen et al., 2018; Heskes et al., 2018). In plants, MSI has been used to target small molecules in different plant species and tissues. Recent examples include *Arabidopsis* leaves (Shroff et al., 2008), Medicago roots (Ye et al., 2013), wheat seeds and stems (Burrell et al., 2006; Robinson et al., 2007), soya leaves (Mullen et al., 2005), rice seeds (Zaima et al., 2010a), different tissues of Eucalyptus species (Hansen et al., 2018; Jakobsen Neilson et al., 2019), and barley roots (Sarabia et al., 2018). Novel metabolites were discovered using Matrix Assisted Laser Desorption Ionization-Mass Spectrometry Imaging (MALDI-MSI) (Gorzolka et al., 2014), further highlighting the benefit of obtaining spatially resolved information.

In this study, we used a combination of four high-end analytical techniques — MALDI-MSI, liquid chromatography-tandem mass Spectrometry (LC-QToF-MS/MS), inductively coupled plasma mass spectrometry (ICP-MS), and micro X-ray fluorescence (μ -XRF), to give new insights into the metabolism of metabolites including lipids, as well as elements including Na^+ and K^+ . These analytical platforms were used to analyze early stages of germination in two barley genotypes that differ in their germination phenology (Mundah-early germinating and Keel-late germinating) under salt stress. Barley was chosen as it is not

only an agriculturally and industrially important crop, but also exhibits a typical cereal germination process, allowing the results to be applied to other cereal crops.

This combined approach provides unique insights into the spatially resolved coordination of metabolic processes that are sequestered among the germinating embryo axis, the lipid rich scutellum, the nutritive endosperm, the digestive enzyme secreting aleurone and the outer pericarp cell layers during the early stages of seed germination. LC-QToF-MS provided information on the distribution of a specific flavonoid in five additional barley genotypes to determine potential roles in response to salt stress. On-tissue analysis using MSI-MS/MS further confirmed the identity of a tentative flavonoid resulting in nine fragment ions. Techniques such as ICP-MS and μ -XRF were used to provide data on element levels that are dynamically mobilized during barley seed germination.

MATERIAL AND METHODS

Plant Material

Seven varieties of barley were selected based on their importance for crop production, commercially relevant traits, and known difference in germination phenology and salinity tolerance (Tavakkoli et al., 2012; Kamboj et al., 2015; Sarabia et al., 2018; Yu et al., 2018). This selection comprised of two Australian barley feed (Mundah and Keel), one Australian food (Hindmarsh), and three malting cultivars (Gairdner, Vlamingh, and Clipper), as well as one North African landrace (Sahara). All seeds were sourced from The University of Adelaide, Australia.

Seed Germination and Sample Collection

The germination of seven barley varieties was compared under normal and saline conditions. Twenty seeds from each variety were sterilized in 70% ethanol for 1 min. They were then rinsed 4–5 times in sterile 18.2 Ω deionized water before being treated with 1.0% (v/v) bleach for 10 min, followed by thorough washing in sterile 18.2 Ω deionized water 6–7 times. Seeds were then imbibed overnight (~16 h) in sterile 18.2 Ω deionized water with continuous aeration before being transferred to 1% agar plates containing a range of salt (NaCl) concentrations: 0 (control), 50, 100, 125, 150, 200, and 250 mM. The plates were then sealed with parafilm, wrapped in aluminium foil and kept at 4°C for 48 h before being transferred to a growth chamber maintained at 17°C constant temperature with no light. Plates were scored for the number of seeds germinating at different times after transfer to the growth chamber (48, 72, 96, and 120 h post-imbibition). The germination percentage was calculated using the following equation: Germination (%) = $G/X \times 100$, where, G = number of seeds germinated and X = number of seeds sown.

Seeds from barley varieties with the highest and lowest germination frequencies under salinity, respectively, were used in a second experiment to produce germinating seeds for MALDI-MSI, ICP-MS, and μ -XRF. The seeds were sterilized as described above. After 16 h of overnight aeration, seeds were harvested for the zero hour time point (0 h) and the remaining seeds were aseptically transferred to petri dishes with and without saline

agar. These dishes were kept in the dark at 4°C for 48 h and then transferred into a growth chamber maintained at a constant temperature of 17°C with no light. Germinating seeds were harvested at 8, 16, 24, 48, or 72 h growth during this time so that seeds at different developmental stages could be analyzed. These seeds were immediately frozen in Super Cryo Embedding Medium (SCEM, Section Lab) using liquid nitrogen and stored at –80°C into 50 ml Conical Centrifuge Tubes (Thermo Fisher Scientific, USA) until subjected to cryosectioning.

Chemicals

Solvents were purchased from Merck Millipore (Bayswater, VIC, Australia), Chemicals including 2,5-dihydroxy benzoic acid (DHB), elemental red phosphorus and Supra pure® nitric acid (70%) were purchased from Sigma-Aldrich (Castle Hill, NSW, Australia). Embedding and freezing supplies including a cryofilm fitting tool set (2.0 cm), embedding container (1.5 cm \times 2.0 cm), embedding medium (SCEM), and cryofilm 2C(9) (2.0 cm in width), were purchased from Section-Lab Co. Ltd. (Tokyo, Japan). Sectioning supplies including Menzel-Gläser Superfrost Ultra Plus Glass slides, Optimal Cutting Temperature (O.C.T.) compound and Feather® C35 tungsten microtome blades were purchased from Grate HDS (Ringwood, Australia). Lysing Matrix Tubes with 0.5 g Lysing Matrix D (1.4 mm ceramic spheres) were purchased from MP Biomedicals (Seven Hills, NSW, Australia). Elemental standards were purchased from PerkinElmer (Melbourne, VIC, Australia). Aqua Regia (concentrated hydrochloric acid and concentrated nitric acid in the ratio 4:1) was prepared fresh immediately prior to use.

MALDI-MSI: Sample Preparation and Imaging

For MALDI-MSI, samples from six time points during germination (**Supplementary Figure S1**) of the two barley varieties were analyzed. Prior to sectioning, frozen seed samples were kept in a Reichert Jung 2800 FRIGOCUT M cryostat at –20°C for 30 min. Later, the frozen sample was attached to a cold specimen disk using Optimal Cutting Temperature (OCT) compound and trimmed with a C35 tungsten microtome blade (Feather®). Longitudinal sections of 14 μ m were stabilized with cryofilm 2C(9) (Section Lab, Japan), attached to a glass slide with electrically conducting double-sided tape according to the Kawamoto's film method with slight modifications (Kawamoto, 2003; Zaima et al., 2010b; Boughton et al., 2016), and immediately freeze dried in a Christ Alpha 1–4 LD plus Freeze Drier (John Morris Scientific, Australia) under vacuum (1.0 mbar) for 20 min. This step was important in order to eliminate water from seed sections and also to facilitate their manipulation and storage at room temperature before further analysis. The sections were chosen from the centre of the seed so the embryo, endosperm and aleurone layer were included in all sections. In order to account for biological variation present, at least three different seeds from each time point and treatment were analyzed. These were treated as technical and biological replicates.

2,5-dihydroxy benzoic acid (DHB) was used as sublimation matrix for positive mode analyses. For this study, the optimal

conditions were 0.11 Torr pressure during sublimation, a hot plate temperature of 140°C, 300 mg DHB as matrix added to the sublimator, and a sublimation period of 4 min.

MS imaging of seed sections was performed on a Bruker Solarix 7T hybrid ESI-MALDI-FT-ICR mass spectrometer equipped with a Smartbeam II laser (Nd : YAG laser 355 nm) operating at 2 kHz, using the “minimum” focus setting and smart walk enabled (Bruker, Bremen, Germany). For every run, laser fluence was optimized to obtain the best signal-to-noise (S/N) ratio. The MS was operated in the positive ionization mode across the mass range m/z 150–2,800. MSI measurements were acquired with a raster of $75 \times 75 \mu\text{m}$. The measurement of the spots was carried out in random order to eliminate influences of measurement order. An external calibration (quadratic equation) was carried out using elemental red phosphorus obtaining a minimum of seven points of calibration over the mass range. A mass accuracy better than 4 ppm was obtained across a tissue section image. Co-registration between each MSI data set and its optical image was performed with flexImaging software v.4.1 (Bruker Daltonics, Bremen, Germany).

MS/MS spectra of selected ions, previously identified through SCiLS Lab (SCiLS Lab 2017a, SCiLS GmbH) analysis were conducted in positive mode using a laser power of 40%. A total of 5,000 laser shots per sample were fired randomly across the relevant tissue region. The isolation window was set to 5 m/z with the collision RF amplitude set to 1,200.0 Vpp and collision voltage set to 20 V.

MALDI-MSI: Analysis, Data Processing, Detection and Annotations

The raw MALDI-MSI datasets were analyzed using flexImaging software v.4.1 (Bruker Daltonics, Bremen, Germany) and normalized to the Root Mean Square (RMS). Any m/z peaks that showed spatial distribution across the tissue sections were manually annotated and used to create a list of spatially distributed potential mass features. The generated list was further analyzed to identify protonated ($[M+H]^+$), sodiated ($[M+Na]^+$) and potassiated ($[M+K]^+$) adducts of the mass features that displayed spatial distribution prior to identification of potential metabolites using accurate mass search against the LIPID MAPS (Fahy et al., 2007) databases by accurate precursor mass (<5 ppm). Further average MS spectra were also exported from flexImaging and imported into SimLipid v.5.50 (Premier Biosoft International, Palo Alto, CA, USA) for accurate mass annotation. Tentative annotations were performed to level two based on accurate mass match and spectral libraries (Sumner et al., 2007). All potential lipid and metabolites detected with multiple adducts were collapsed into single ID for identification and analysis purposes.

Annotated m/z features were imported into SCiLS Lab (SCiLS Lab 2017a, SCiLS GmbH; (<http://scils.de/>; Bremen, Germany) software from flexImaging and further used to identify discriminant features between control sections and treatments. Sub regions were created to analyze discriminant lipids in embryo and endosperm regions. Normalization of data was performed according to the Total Ion Count (TIC) of all spectra

prior to data analysis. Next, discriminative m/z values were identified from all individual spectra with a minimal interval width of ± 5 mDa between control and salt-treated sections. A computer-generated diagram known as a Receiver Operating Characteristic (ROC) curve was used to quantify how well the m/z values discriminated between two treatments (control and salt). Strict ROC Area Under the Curve (AUC) values of ≥ 0.8 and ≤ 0.2 were set as an additional criterion to the Student's t -test p -value of < 0.005 for peaks considered as statistically significant for salt vs control and control vs salt respectively (Rausser et al., 2010). The Intensity Box Plot chart was generated which depicts intensities of a given m/z interval filtered by the visible regions through their quartiles.

Peak lists were then exported and reimported into flexImaging to verify and confirm the spatial distribution of identified discriminant peaks. After analysis, a curated list of spatially distributed peaks was generated for control vs salt-treated images. For MS/MS analysis, raw data was analyzed on Bruker Compass DataAnalysis v5.0 using the deconvolution and manual annotation tools.

ICP-MS: Sample Preparation and Analysis

Dry seeds were ground to a fine powder using a Tube Mill control (IKA Mills, Selangor, Malaysia) at 25,000 rpm ($30,500 \times g$) for 4 min. The ground samples were accurately weighed (20 mg) into 10 ml Falcon tubes (Sarstedt, Australia). These samples were then digested in 300 μl acid solution (Aqua Regia) by heating for 1.5 h at 80°C. After digestion and cooling to room temperature, deionized water was added to make the final volume to 10 ml. The tubes were then centrifuged at 5,000 rpm ($6,100 \times g$) for 5 min. The supernatant was used for ICP-MS analysis.

A modified method from Callahan et al. (2016) was used for this study. An ICP-MS instrument (NexION 350X, PerkinElmer, USA) was used to measure the concentration of the following elements: P, S, K, Ca, Mg, Mn, Co, Ni, Cu, Zn, and Fe. The internal standards Sc (200 ppb) and Rh (20 ppb) in 1% Aqua Regia were used for correction. The internal standard was mixed with the sample stream *via* a T-piece in a 1:1 ratio giving a final acid concentration of 2% at the source. Two sets of calibration standards were prepared. Set 1 contained P, Na, K, and Ca at 500, 1,000, and 5,000 ppb. The second set contained the other elements at 0.1, 1, 10, 50, 100, and 500 ppb. The mass spectrometer was operated in kinetic energy discrimination mode (KED) with 50 ms dwell times, 20 sweeps, one reading and three replicates. The plasma source conditions were: nebulizer gas flow 1.02 L min^{-1} , auxiliary gas flow 1.2 L min^{-1} , plasma gas flow 15 L min^{-1} , ICP RF power 1,500 W.

Data were processed using Syngistix (PerkinElmer) software. Signal responses were normalized to the rhodium (Rh) internal standard, sample weights, and dilution factors.

μ -X-ray Fluorescence Analysis

An M4 Tornado μ -XRF spectrometer (Bruker, Bremen, Germany) equipped with a Rh anode side window X-ray tube with a polycapillary lens was used to obtain elemental distribution maps in this study. The step size was set to $40 \mu\text{m}$ (for whole seed)

combined with high excitation intensity. Direct analysis on seed sections prepared similarly as described in Section 3.4 (without matrix application). This was achieved by placing the sections on the μ -XRF platform under 20 mbar vacuum, 50 kV voltage, and an anode current of 199 μ A conditions (Sarabia et al., 2018). Detailed summary on the instrumental conditions and operating parameters used to analyze barley seeds in this study is given in **Table S1**. Spectra acquisition and 2D elemental maps were obtained using the Esprit software (Bruker, Berlin, Germany).

LC-QToF-MS: Sample Preparation and Analysis for Flavonoid Detection

Seeds were germinated and harvested as described above (Section 2.2) followed by immediate freeze drying in a Christ Alpha 1–4 LD plus Freeze Drier (John Morris Scientific, Australia) under vacuum (1.0 mbar) for 48 h. This facilitated fine powdering of seeds prior to extraction. A modified method from (Routaboul et al., 2006) was used for extracting flavonoids. Briefly, flavonoid extracts were obtained from 30 mg dried seed powder. Powder was transferred into cryo-mill tubes along with 1 ml acetonitrile/water (75:25; v/v) containing the following internal standards, $^{13}\text{C}_6$ -Sorbitol (0.02 mg/ml) and $^{13}\text{C}_5^{15}\text{N}$ -Valine (0.02 mg/ml). The samples were homogenized using a cryomill (Bertin Technologies; 6,000 rpm (7,320 \times g) for 3 repetitions at 45 s with a 45 s interval between replicates) at -10°C . The mixture was then sonicated for 20 min followed by centrifugation at 13,000 rpm (15,900 \times g) for 10 min at room temperature. The supernatant was separated and transferred into a fresh 2 ml tube. The remaining pellet was further extracted with 1 ml acetonitrile/water (75:25; v/v) on a thermomixer, shaking overnight (500 rpm, 4°C). The solution was centrifuged, and the two extracts were pooled, dried in a SpeedVac (Martin Christ Gefriertrocknungsanlagen GmbH, Germany) without heating.

Dried flavonoid extracts were resuspended in 200 μ l acetonitrile/water (75:25; v/v), and 5 μ l (for MS analysis) or 10 μ l (for MS/MS analysis) aliquots were injected onto a Poroshell 120, HILIC-Z, 2.1×100 mm (2.7 μ m particle size) column (Agilent Technologies, Santa Clara, CA, USA) at 25°C , and analyzed on a LC system coupled to an Agilent Technologies (Santa Clara, CA, USA) 6545B ESI-QToF-MS. Flavonoids were eluted at 0.250 ml/min over 18 min with a binary gradient of ACN-water (90:10, v/v) containing 20 mM ammonium formate and formic acid and water containing 20 mM ammonium formate and formic acid as described by Routaboul et al. (2006). MS data were acquired with an ESI source in positive ion mode. The MS data presented corresponds to ten pooled biological replicates for each treatment group. MS/MS was performed to obtain the MS2 spectra of the identified compound.

LC-QToF-MS: Data Processing and Analysis

LC-MS data was processed using MassHunter Qualitative Analysis and MassHunter Quantitative Workstation v.B.08.00 software (Agilent Technologies, Santa Clara, CA, USA) for flavonoid identification. Peak spacing tolerance was set to 0.5 m/z with a retention time window of 0.1 s. The acquired

data was exported and normalized with sample weight and log transformed. Univariate statistical analysis tests (Student's t -test) was carried out using Minitab software. X-fold changes were determined between control and treated samples of seven genotypes.

RESULTS

Overview of Experimental Workflow

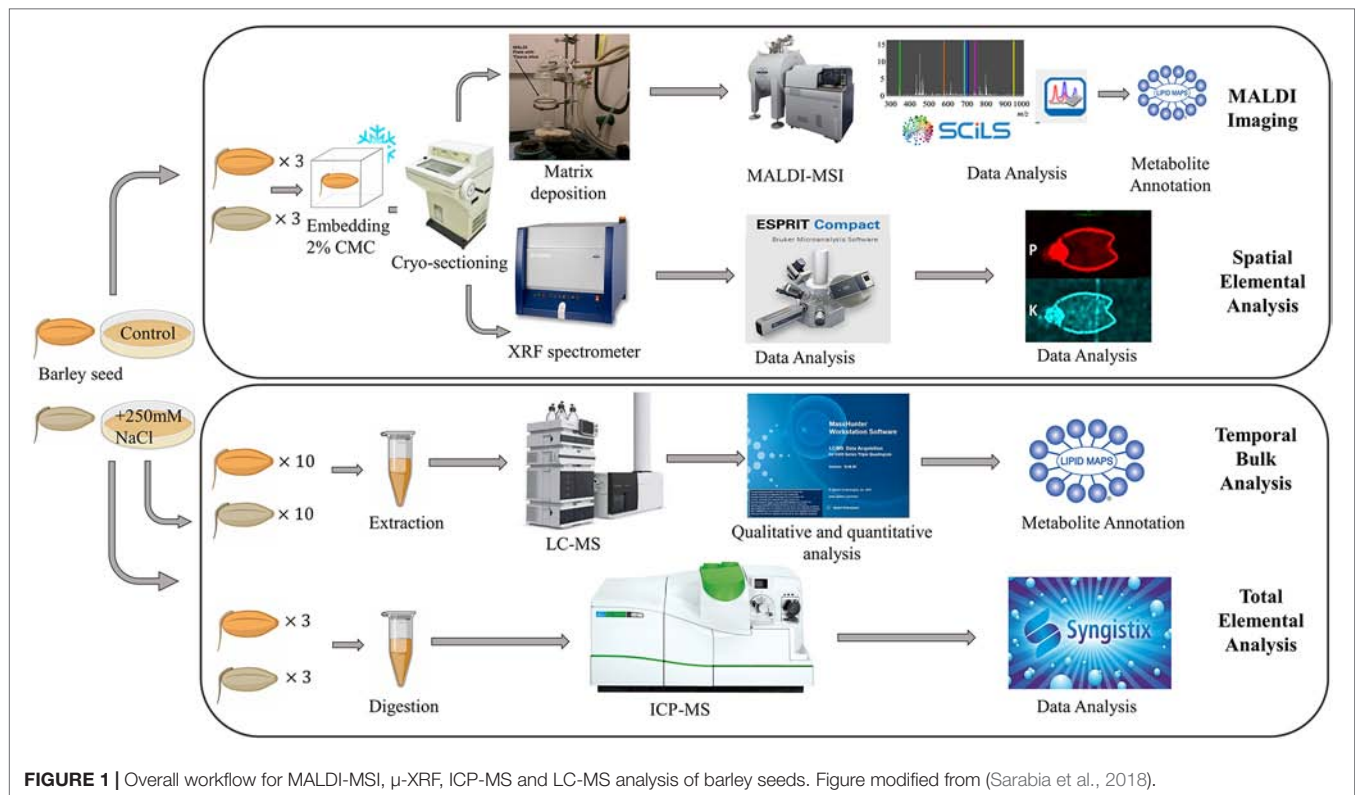
A schematic of the experimental workflow is provided in **Figure 1**. MALDI-MSI analysis was performed for seeds grown under control and saline conditions and harvested at six different time points, namely 0, 8, 16, 24, 48, and 72 h post imbibition (**Figure S1**). Sublimation of sectioned tissue with DHB matrix was used for MALDI-MSI positive mode ionization to image the spatial distribution of a wide range of lipids (m/z 150–2,800) in the seed sections. ICP-MS (for total elemental analysis) and μ -XRF spectrometry (to obtain the spatial information of targeted elements) was also performed for three time points (0, 24, and 48 h post imbibition).

Germination Assay: Germination and Seedling Growth Under Salt Stress

In the initial germination assay, seeds from seven barley genotypes (Vlamingh, Mundah, Clipper, Sahara, Hindmarsh, Gairdner, and Keel) were germinated in petri dishes containing different salt concentrations (0–250 mM NaCl) and observed for germination from 0–120 h post imbibition (**Table S2**). Seeds were considered to be germinated once emergence of the cotyledon was visible. There was a clear difference in the germination frequency of the seeds germinating in the presence of salt. Keel showed the greatest inhibition (only 65% germinated at highest salt concentration) while all Mundah and Vlamingh seeds germinated. Other genotypes showed germination frequencies between 75 and 90%. The most tolerant genotypes were viable under saline conditions for at least 7 days. Based on this assay, two Australian barley feed varieties were chosen to investigate the metabolic changes that occur during germination under salinity: Mundah, a very early spring variety from Western Australia with very high early vigor (highest germination rate under salinity), and Keel, an early spring variety from South Australia with moderate early vigor (lowest germination rate under salinity) (Long et al., 2003).

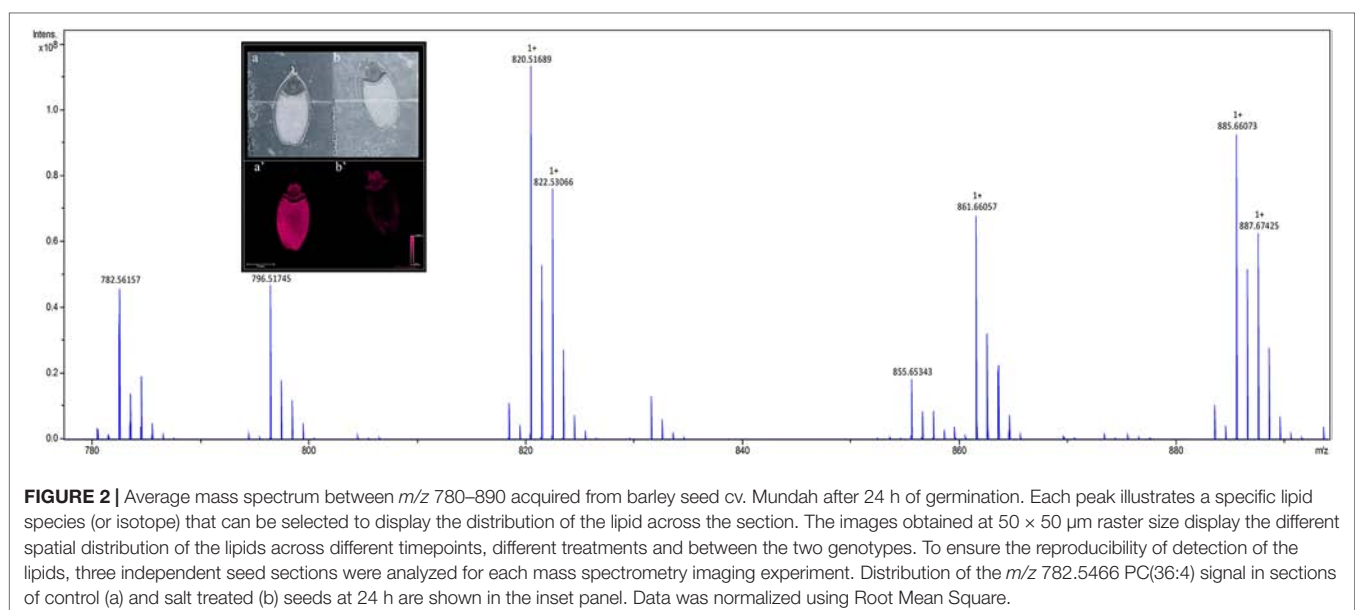
MALDI-MSI of Lipids in Barley Seeds During Germination

To identify lipids and other small molecules, the distribution of these compounds at all post imbibition time points in control and saline conditions were compared in both genotypes. Lipids within longitudinal sections of mature barley seeds were profiled by MALDI-MSI at six time points post imbibition (**Figure S1**). Signals derived from the DHB matrix were excluded from further analyses. A total of 774 peaks in the average MS spectra were found to have a differential spatial distribution in the samples. An example of an average mass spectrum between m/z 780–900 acquired from a Mundah seed after 24 h post imbibition



is presented in **Figure 2**, illustrating a specific peak that can be selected to display the distribution of a corresponding lipid across the section. Many low abundance lipids were not reproducibly detected across all sections due to analytical and biological variations and thus, only the presence and tissue distribution of the tentatively identified lipids observed across three biological replicates are discussed (**Figure S2**).

All peaks showing spatial distribution in the seed sections were selected from flexImaging (**Tables S3 A–D**). Mass precursor ion search (<5 ppm) of the LIPID MAPS database (Fahy et al., 2007) was performed to provide tentative annotations of the features detected and resulted in between 246 to 319 tentatively assigned mass features for each sample. Some of these mass features corresponded to $[M+H]^+$, $[M+Na]^+$, and $[M+K]^+$ ion adducts of



the same lipid and were observed in all samples. Two hundred and sixty-nine lipid species were identified in both the Mundah control and salt treated seeds. Three hundred and nineteen lipid species were identified in Keel control seeds but only two hundred and forty-six in salt treated Keel seeds. Lipids were annotated to the sum level only using the Liebisch notation (Liebisch et al., 2013). In-source fragmentation of more complex lipid species, such as Phosphatidylcholine (PC), during detection, may lead to incorrect identification of fragments belonging to other lipid species such as Phosphatidic Acid (PA), Fatty acyls (FA) and Lysophosphatidylcholines (LPC) (Wang, 1999) and this was taken into account during annotation of the lipid species. For example, it was demonstrated that PC, PA, and LPC species had different spatial distribution in seeds (**Figure 3**) suggesting that intact lipid species are being detected and not fragments of other lipids.

Tentatively annotated lipid species with several ion adducts, but the same masses, were collapsed into a single ID. It was noted that the returned matches were filtered to remove tentative annotations from mammalian origins. As a result, 234 lipid species were tentatively identified in Mundah control (**Table S4A**), 202 in salt treated Mundah seeds (**Table S4B**), 155 in Keel control (**Table S4C**) and 145 in salt treated Keel seeds (**Table S4D**). The tentatively annotated lipids were categorized into seven major lipid classes: Fatty acyls (FA), Glycerolipids (GL), Glycerophospholipids (GP), Sphingolipids (SP), Prenol lipids (PR), Sterol lipids (ST), and Polyketides (PK). The percentage of each lipid class in both genotypes is given in **Table 1**. This was solely based on the number of lipid species matched with the LIPID MAPS database. The distribution of lipid classes at all timepoints in both genotypes are shown in **Supplementary Figure S3**. For both genotypes, the number of lipids identified is smaller in the germinated seeds in the presence of salt compared to the control, but in Mundah, the difference between the two treatments is much smaller. The number of lipids identified in each class decreased in Keel after salt treatment, whereas in Mundah, only GL, GP, and ST decreased.

TABLE 1 | Percentage of individual lipid classes in two treatments of Mundah and Keel. The values are total percentages of annotated lipids at all timepoints.

Lipid classes	Mundah (%)		Keel (%)	
	Control	Salt	Control	Salt
Fatty acyls	18.70	20.00	26.60	26.62
Glycerolipids	14.07	13.85	24.68	23.55
Glycerophospholipids	49.81	44.92	26.28	34.13
Polyketides	7.96	10.15	7.69	6.48
Prenol lipids	1.85	2.15	5.13	3.07
Saccharolipids	0.56	0.00	0.00	0.00
Sphingolipids	3.89	4.31	3.53	2.39
Sterol lipids	3.15	4.62	6.09	3.75

In Mundah seeds, GP was the major class of lipid altered (43% in Mundah control and 40.6% in Mundah salt). In the GP class, PA, Phosphatidylethanolamine (PE), Phosphoglycerols (PG), and PC were tentatively identified as the major lipid subclasses. The GL class decreased in saline conditions but in control conditions, TAGs were detected in high number. The percentage of ST and GP decreased under salinity, but the sterol/phospholipid ratio was unaffected by salt (0.25% in control and 0.24% in salt).

GP is also the major class altered in Keel seeds (26.3% in Keel control and 34.1% in Keel salt). It is worth noting that an increased number of GP lipids in Keel salt treated seeds was detected likely due to different adducts detected for the same lipid. For example, in the GP class, LPC (16:0), LPC (18:0), and PC (36:5) were found with $[M+H]^+$ and $[M+K]^+$ ion adducts and PC (34:2) with $[M+Na]^+$ and $[M+K]^+$ ion adducts. The percentage of ST decreased significantly under salt stress in Keel with a slight decrease in GP under salt stress and hence the sterol/phospholipid ratio was affected by salt in Keel seeds (0.25% in control and 0.12% in salt). It is noted that PC lipid species of 36:n family have been found to be differentially distributed at all timepoints in both genotypes under control and salt conditions.

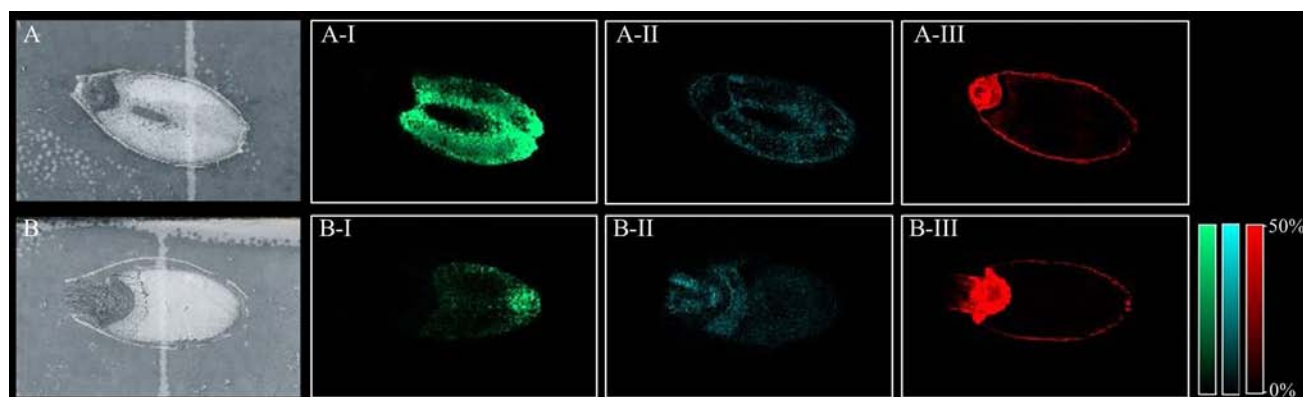


FIGURE 3 | Different lipid species displaying different spatial distribution. **(A)** scanned image of Mundah seed at 48 h in control conditions; **(B)** scanned image of Mundah seed at 48 h in salt conditions; A-I and B-I: m/z 496.338 LPC(16:0) showing spatial distribution only in endosperm of both seeds with different intensities in control and salt conditions. A-II and B-II: m/z 757.574 PA(40:2) showing spatial distribution in entire seed with different intensities in control and salt treated seed; A-III and B-III: m/z 820.530 PC(36:4) showing spatial distribution in the aleurone layer and embryo of control and salt treated seeds. Intensity scale for both images was set to 0–50%. Data was normalized using Root Mean Square.

Gorzolka et al. (2016) performed MSI to localize metabolites during the first seven days of germination in barley seeds and showed the distribution of 11 lipids in barley seeds at different stages of germination and different regions of seeds. In our study, we confirmed the presence of 10 out of 11 lipids from Gorzolka et al. (2016) in Mundah control, 9 in Mundah salt, 4 in Keel control and 9 in Keel salt treated seeds (listed in **Table S5**). Time-dependent changes were also observed in the lipid profile of the endosperm early in germination in Gorzolka et al. (2016). Mono- and diacyl PCs (m/z 496, 520, and 758) were highly abundant, and these lipids changed their distribution at later germination stages (Gorzolka et al., 2016). In the current study, the m/z 496 and 758 ions could be detected in both genotypes under both treatments. However, they were not consistently observed across all time points. In contrast, the m/z 520 ion was only detected in salt treated Keel seeds. The distribution pattern for these three PCs is given in **Table S6**.

Lipids Detected Immediately After Imbibition and After 8 h of Salt Stress

Comparison between the annotated lipid classes at 0 h and 8 h post-imbibition in control Mundah seed sections showed 26 lipids species present at both time points whereas only 16 lipids species were found in common when 0-hour and 8-hour salt treated seeds were compared (**Tables S7A, B**). In Keel, 30 lipids species were detected in both 0-hour and 8-hour control sections whereas 20 lipids species were found between 0-hour control and 8-hour salt treated seed sections (**Tables S7C, D**). After 8 h of salt stress, 65 lipid species were differentially accumulated compared to the 8 h post imbibition in untreated seeds of Mundah (**Table S8**). The major lipid class detected in 8 h salt treated seeds in Mundah was GP. In contrast, 38 lipids were detected after 8 h of salt stress when compared to the 8 h untreated seeds in Keel (**Table S9**). The major lipid class detected in Keel after exposure to salinity stress was GL.

Spatial Segmentation of the MALDI-MSI Data Using an Unsupervised Clustering Approach Using SCiLS Lab

Relatively few lipid species changed their spatial distribution in Keel compared to Mundah seeds. In Mundah, out of 225 lipid species that were found to be spatially distributed in salt treated seeds at all time points, 71 (31.5%) were salinity treatment specific. In Keel, out of 163 lipid species spatially distributed in salt seeds at all time points, 48 (29.4%) were specific to salt stress based on LIPID MAPS annotations.

SCiLS Lab was used to determine the discriminative features for both treatments in Mundah and Keel. A univariate measure was used to assess the quality of discrimination for imported m/z values for masses created in flexImaging, which quantified how well the m/z values discriminated between two treatments. A threshold for Area Under ROC (Receiver Operating Characteristic) Curve (AUC) was used to determine the discriminating masses in both treatments for both genotypes. A perfect discrimination was accepted where AUC value was equal to or close to 1 for salt treated seeds and close to or equal

to zero for control seeds. The threshold for this analysis was set to 0.8 for salt treated seeds and 0.2 for control seeds. Subregions were created in SCiLS Lab to find m/z values that discriminate between the two treatments for embryo and endosperm in both genotypes. An example of a subregion created for Mundah seeds (m/z 615.494) and its corresponding ROC plot for DAG (34:2) with AUC value of 0.087 is given in **Figure S4**. The intensity box plot chart that describes a single m/z interval and a single region was also used.

Further discriminant lipids were detected at all timepoints using the above-mentioned threshold values in both genotypes. Two subregions were created — embryo and endosperm. In Mundah embryo, lipids discriminant to control seeds (AUC close to zero) increased in number over time whereas in Keel embryo, the number of discriminant lipids with AUC close to zero decreased over time. In Mundah endosperm, 45 lipids were found to be discriminant between control and salt treated seeds with only nine lipids showing AUC value between 0.8–1.0. In Keel endosperm, only four lipids were discriminant with AUC values between 0–0.2 (**Table S10**).

In Mundah embryos, a tentative flavonoid, m/z 365.102 [$M+H$]⁺ (AUC for Mundah 16 h, 0.867; Mundah 24 h, 0.949; Mundah 48 h, 0.907) was found to be strongly discriminant at 16, 24, and 48 h (**Figures 4A–C, A'–C'**) whereas the same mass was observed in Keel embryos at 72 h only (AUC 0.946) (**Figures 4D, D'**). Signal intensity was scaled to optimize visualisation of the respective ion. The intensity box plot and ROC plot for the above mentioned timepoints shows the discrimination capabilities of the given m/z for two (control and salt) regions in Mundah at 16, 24, and 48 h (**Figure S5A**) and Keel at 72 h respectively (**Figure S5B**). No common lipids were found in Mundah embryos over all time points. In Keel embryos, diacylglycerols (m/z 602.523) and a fatty alcohol (m/z 576.508) discriminate between control and salt treatments at 16 and 24 h with an AUC of 0.19 and 0.13 respectively. In Mundah endosperms, it is worth noting that fatty acyls (m/z 339.288, [$M+Na$]⁺) were found in salt treated seeds at 48 h with an AUC 0.88 whereas at 72 h, they were found intensively in control seeds with AUC value of 0.17. No discriminant lipids were found in Keel endosperms. Although several lipid species were found distinct at 8 h salt as compared to 8 h control as mentioned in Section 3.4, no discriminant lipids between these two timepoints in both genotypes were found.

Confirmation of Tentative Flavonoid Using MS/MS MALDI-MS/MS

To probe the identity of the tentative flavonoid (m/z 365.102 [$M+H$]⁺), on-tissue MALDI-MS/MS spectra was generated and identified fragments were recorded. A 5 m/z window centred around m/z 365.1 was isolated with no collision voltage applied. **Figure 5** shows the two distinct peaks with very close mass obtained from MALDI-MS, m/z 365.1024 was matched in the LipidMaps and METLIN (<http://metlin.scripps.edu>) databases to a flavonoid and m/z 365.1058 which was assigned as a dihexose sugar (Calvano et al., 2017). Subsequent MS/MS using the same isolation window and application of 20 V

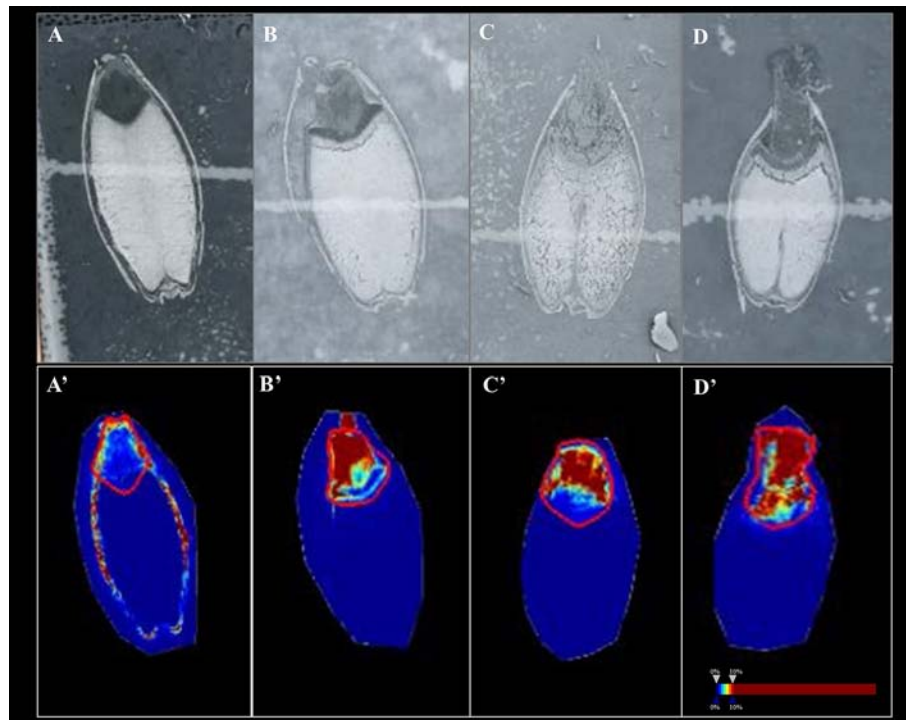


FIGURE 4 | A flavonoid (m/z 365.102) with AUC close to 0.8 at 16, 24 and 48 h in Mundah embryos and at 72 h in Keel embryos. **(A, B and C)** scanned images of Mundah seeds at 16, 24 and 48 h of salt treatment; **(D)**: scanned image of Keel seed at 72 h of salt treatment; **(A', B' and C')**: Embryo region showing high intensity of the flavonoid in Mundah salt treated seeds at 24 and 48 h; **(D')**: Embryo region showing high intensity of the flavonoid in Keel salt treated seed at 72 h. Intensity scale was set to 0–10% for all seeds. Data was normalized using Root Mean Square.

collision energy led to observation of nine product ions (**Figure S6**) that were matched with the *in silico* predicted spectra of a flavonoid, Gancaonin F (METLIN database) (Smith et al., 2005), corresponding ions associated with the fragmentation of a dihexose sugar were also observed.

LC-MS/MS

LC-MS/MS was performed on seed extracts to provide further evidence for annotation of the possible flavonoid and dihexose sugar. Both MS and MS/MS were conducted with the protonated precursor ion m/z 365.1 isolated for MS/MS. The extracted compound chromatogram (ECC) at the MS level showed two separate peaks with retention time of 6.71 and 4.49 min. MS/MS using 40 eV and 20 eV respectively showed distinct differences in the corresponding spectra between the two compounds. Inset **Figure 5A** displays the mass spectrum for one compound with seven fragments, of which three fragments (m/z 347, 323, and 307) match closely with *in silico* predicted spectra of Gancaonin F (METLIN database) and MALDI-MS/MS data obtained from on-tissue analysis in this study. Inset **Figure 5B** shows the mass spectrum of a compound with two fragments (m/z 203 and 185) that match exactly to dihexose sugar fragments (Calvano et al., 2017). The mass spectrum along with the database search showed a difference of less than 4 ppm in experimental and database mass identified using LC-MS/MS. **Table 2** shows the observed fragments of the tentative flavonoid detected using MALDI-MS/

MS and LC-MS/MS that matched with the *in silico* predicted spectra of Gancaonin F.

Level of Tentative Flavonoid Increases in Response to Salinity Stress in Four Out of Seven Barley Genotypes

The response of the tentative flavonoid was further elucidated in seven barley genotypes (Mundah, Keel, Vlamingh, Gairdner, Hindmarsh, Clipper, and Sahara) at five timepoints (8, 16, 24, 48, or 72 h) in control and saline conditions (250 mM NaCl) using LC-QToF-MS. The aim was to understand the differential expression of the detected tentative flavonoid in germinating seeds of a barley variety panel consisting of a small but representative set of six barley genotypes. **Figure S7** shows the differential change in the tentative flavonoid over time in seven genotypes. Hindmarsh, Clipper, and Gairdner showed no significant difference over time in both conditions. In Sahara, the highest increase was shown at 16 h under saline conditions. In Keel, a significant increase was observed at 72 h under saline conditions as compared to control seeds. A similar pattern was observed in Vlamingh where significant increase in the tentative flavonoid was seen at 72 h in salt treated seeds as compared to control seeds. In Mundah salt treated seeds, there was a decrease in the tentative flavonoid at 72 h. **Table S11** shows fold changes in signal response of the flavonoid for seven genotypes over time.

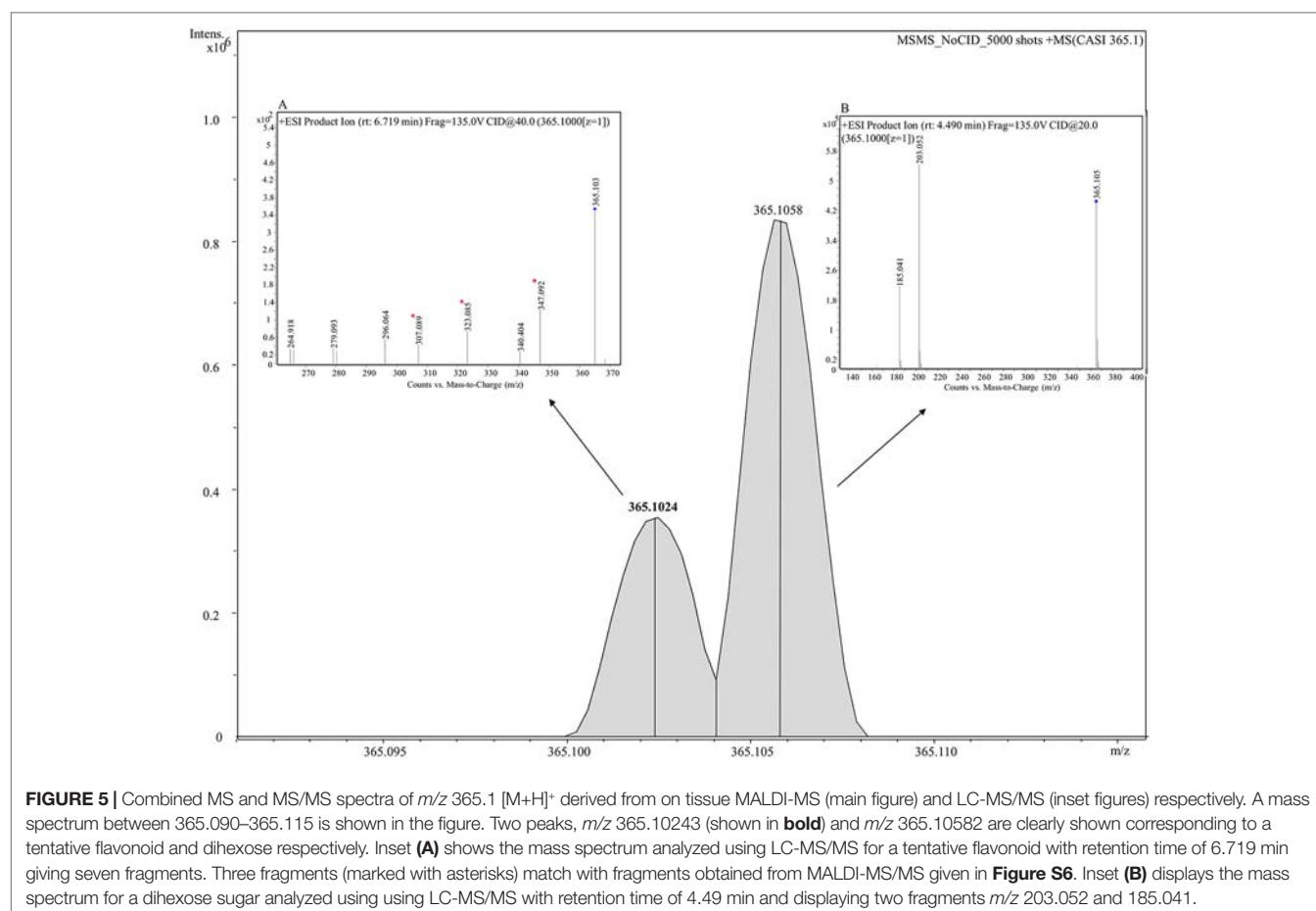


TABLE 2 | Theoretical and observed m/z for the tentative flavonoid, Gancaonin F from MALDI-MS/MS and LC-MS/MS.

<i>In silico</i> predicted fragments of Gancaonin F	Observed m/z by MALDI-MS/MS	MALDI-MSI mass error in ppm	Observed m/z by LC-MS/MS	Mass error in Da
347.091	347.092	2.9	347.092	<0.001, <0.001*
337.107	337.103	11.9	—	0.004
335.091	335.092	2.9	—	0.001
333.076	333.076	<0.001	—	<0.001
323.055	323.055	<0.001	323.085	<0.001, 0.03*
321.112	—	—	—	—
309.039	309.040	3.2	—	0.001
307.060	307.060	<0.001	307.089	<0.001, 0.029*
297.039	297.040	3.4	—	0.001
281.044	281.045	3.5	—	0.001

In silico fragments are obtained from METLIN (<http://metlin.scripps.edu>).

*mass error observed from LC-MS/MS.

Under saline conditions, among all genotypes, Vlamingsh (+6.3-fold) and Keel (+14.04-fold) show the highest increase after 72 h salt stress ($p < 0.001$). However, in Mundah the highest increase was observed at 48 h (+9.0-fold) after salt stress with small decrease at 72 h (+8.9-fold) after salt stress.

Elemental Concentrations in Seeds

The elemental composition of total barley seeds was analyzed to determine the effect of 24 and 48 h of salt stress (250 mM NaCl)

on the content of eight macro- and micronutrients of both barley genotypes. **Table 3** summarizes the concentrations of elements in control and salt treated seeds and **Figure 6** shows the change in Na^+ concentrations in both genotypes after 24 h in control and salt treated seeds. Of all the elements detected, only Na^+ content was significantly altered ($p < 0.001$) after exposure to salt. In Mundah, Na^+ concentrations were 0.5 ± 0.12 mg g^{-1} in control and 3.1 ± 0.48 mg g^{-1} after salt treatment for 24 h and 0.5 ± 0.06 mg g^{-1} in control and 4.7 ± 0.06 mg g^{-1} after salt treatment for 48

TABLE 3 | ICP-MS of whole barley seed (Mundah and Keel) and its main tissue fractions (mg g^{-1} dry weight). Each tissue type was weighed and analyzed separately by ICP-MS.

Element	Mundah (24 h)		Mundah (48 h)		Keel (24 h)		Keel (48 h)	
	Control	Salt	Control	Salt	Control	Salt	Control	Salt
P 31	12.41 \pm 1.69	12.89 \pm 1.99	12.64 \pm 2.89	14.29 \pm 2.13	11.07 \pm 1.9	12.17 \pm 2.6	12.54 \pm 4.43	13.34 \pm 2.12
*Na 23	0.53 \pm 0.12	3.12 \pm 0.48	0.50 \pm 0.06	4.71 \pm 0.06	0.44 \pm 0.02	2.74 \pm 0.21	0.45 \pm 0.06	4.16 \pm 0.11
Mg 24	2.95 \pm 0.22	2.72 \pm 0.74	2.7 \pm 0.61	2.94 \pm 0.62	1.81 \pm 0.5	2.21 \pm 0.61	2.17 \pm 0.95	2.36 \pm 0.63
K 39	8.63 \pm 1.24	8.23 \pm 1.91	8.3 \pm 2.04	8.05 \pm 0.8	7.12 \pm 1.18	7.01 \pm 0.51	7.71 \pm 2.49	7.58 \pm 0.9
Ca 44	1.16 \pm 0.08	0.76 \pm 0.24	1.06 \pm 0.3	0.85 \pm 0.22	0.79 \pm 0.22	0.7 \pm 0.16	0.88 \pm 0.39	0.79 \pm 0.24
Mn 55	0.04 \pm 0.01	0.04 \pm 0.01	0.04 \pm 0.01	0.04 \pm 0.01	0.03 \pm 0.01	0.03 \pm 0.01	0.03 \pm 0.01	0.03 \pm 0.01
Fe-3 57	0.06 \pm 0.01	0.06 \pm 0.02	0.06 \pm 0.02	0.06 \pm 0.02	0.04 \pm 0.01	0.05 \pm 0.02	0.05 \pm 0.03	0.06 \pm 0.02
Zn 66	0.06 \pm 0.01	0.06 \pm 0.01	0.06 \pm 0.02	0.07 \pm 0.02	0.05 \pm 0.01	0.06 \pm 0.03	0.05 \pm 0.03	0.07 \pm 0.02

Data points represent as mean \pm standard error, $n = 3$. Scale in $\text{mg g}^{-1}/\text{DW}$.

*Values with significant differences ($p < 0.001$).

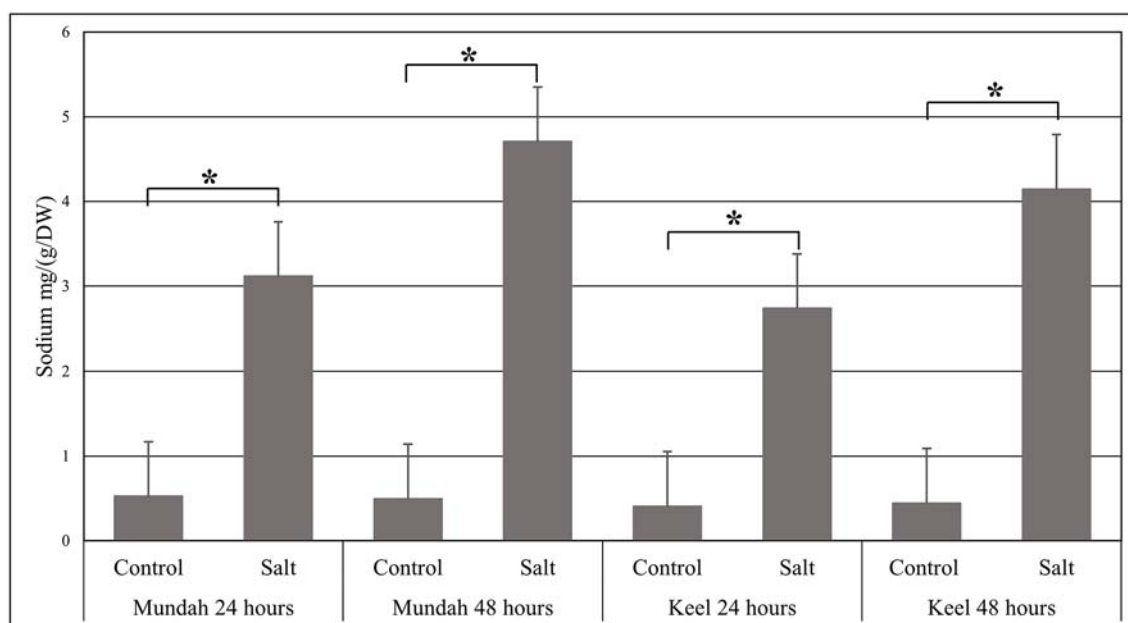


FIGURE 6 | Change in the sodium concentrations identified by ICP-MS in Mundah and Keel seeds after 24 h and 48 h in control and salt treated conditions. Concentration of sodium $\text{mg g}^{-1}/\text{dry weight}$ ($n = 4$). Scales over the bar graph indicates SDs. Asterisks on the top of each bar highlight the significance levels between two treatments at individual timepoints (* $P < 0.001$).

h. In Keel, Na^+ concentrations were $0.4 \pm 0.02 \text{ mg g}^{-1}$ in control and $2.7 \pm 0.21 \text{ mg g}^{-1}$ after salt treatment for 24 h and $0.4 \pm 0.06 \text{ mg g}^{-1}$ in control and $4.2 \pm 0.11 \text{ mg g}^{-1}$ after salt treatment for 48 h. This indicated a significant increase in Na^+ content in both Mundah (+9.4-fold) and Keel (+10.1-fold) after 48 h salt stress.

The K^+ content in seeds decreased slightly after salt treatment but this was not statistically significant. None of the other six elements (P, Mg, Ca, Mn, Fe, and Zn) showed a statistically significant difference in seeds after salt stress as shown in **Table 3**. Since the overall concentration of most elements did not change during germination or in response to salt treatment we examined whether the treatments caused changes in the localization of the elements (S, P, K, and Cl) using $\mu\text{-XRF}$ analysis on seed sections from time point 24 and 48 h. In Mundah seeds, S accumulated in the entire seed irrespective of treatment. P was distributed in

all seeds at moderate concentrations and was highly abundant in the aleurone layer. K was found in higher concentrations in control seeds than in salt treated seeds at all time points, with high abundance observed in the aleurone layer. These results complement the data obtained from ICP-MS. We were unable to observe Na^+ due to instrumental limitations. However, Cl^- was used as an indirect measure of the Na^+ distribution in seeds (Sarabia et al., 2018). Cl^- was highly abundant in salt treated seed sections with a uniform distribution correlating with the increased Na^+ content as detected by ICP-MS (**Figure S8**). In Keel, S accumulated across the entire seed similarly to Mundah but was found to be more abundant in sections from salt-treated seeds obtained at 0 and 48 h post imbibition. P was also found to be highly accumulated in the aleurone layer of all sections and with lower abundance in the endosperm. K^+ was seen abundantly

in samples from the 0 h time point followed by a less intense distribution over time in both treatment. Cl^- was prominently abundant in salt treated seed sections correlating with the increased Na^+ content detected by ICP-MS data (Figure S9).

DISCUSSION

In recent decades, salinization of agricultural land has resulted in lower crop productivity. Germination of seeds is directly affected by salinity either by the resulting osmotic stress or by toxic effects of the sodium and chloride ions themselves. In particular, salinity perturbs plant hormone balance (Khan and Rizvi, 1994) and reduces the utilization of seed reserves (Ahmad, 1992). In the past, metabolomics studies have mainly focused on the germination processes in rice and very little has been published on barley or on the effects of saline conditions on the germination process. In addition, these published studies used techniques that do not provide spatial information on metabolic processes. Gorzolka et al. (2016) previously studied the changes in metabolite profiles in barley seeds after 5 days of germination using MALDI-MSI, but changes in the metabolite profiles of barley seeds during the early stages of germination (0 to 72 h post imbibition) were not explored.

The process of uptake of water by a mature dry seed, which occurs in the initial hour of reintroduction of water to the dry seed, is triphasic (Bewley (1997) and results in the resumption of metabolic activity. These metabolic changes are extremely important to successful germination and, as such, it is vital to document and understand what changes occur (Bewley and Black, 1994; Bewley, 1997; Bewley, 2001). We first conducted a germination assay to establish the germination efficiency of seven barley genotypes under salinity stress. From these results, we selected two genotypes with contrasting germination phenology and salinity tolerance for this study: Mundah [early germinating and salinity tolerant, in agreement with Cao et al. (2017) and Keel late germinating and sensitive to salinity].

Following the germination test, ICP-MS and μ -XRF were combined to profile the levels and measure the spatial distribution of elements across six time points in the two selected barley genotypes (Mundah and Keel) under control and saline conditions. Munns and Tester (2008) reported that salt stress results in the accumulation of Na^+ in plants and results in an adverse effect on K^+ concentration which causes detrimental effects to plant growth and development. Various physiological studies have demonstrated that Na^+ toxicity is caused due to the ability of Na^+ to compete for K^+ binding sites to disrupt K^+ homeostasis (Volkov and Amtmann, 2006; Shabala and Cuin, 2008; Hasegawa, 2013). Hence, to avoid an ion homeostasis disorder under saline conditions, plant cells need to maintain a low Na^+ concentration and a concurrent high K^+ concentration in the cytosol, where enzymes for metabolism are located (Zhu, 2003; Pardo et al., 2006). In this study, we observed a substantial increase in Na^+ content and a non-significant decrease in K^+ content in germinating seeds of both genotypes under saline conditions.

It is apparent from this result that accumulation of excessive amounts of Na^+ in germinating seeds are likely to be responsible

for inhibition of germination under salt stress in Keel as compared to Mundah which was still able to germinate under these conditions. This view is supported by a previous report where increased accumulation of Na^+ in wheat seeds under salinity stress inhibited the rate of germination (Begum et al., 1992). The K^+/Na^+ ratio decreased to greater extent in both Mundah and Keel after 48 h compared to 24 h salt stress. This result is in alignment with previously reported reduction of the K^+/Na^+ ratio in barley roots after salt stress (Shelden and Roessner, 2013). Salinity tolerance also depends on limiting Na^+ accumulation and maintaining K^+ content in the cytosol in order to achieve the preservation of ion homeostasis (Flowers et al., 1977; Hasegawa et al., 2000). Cl^- was observed to increase but no change in K^+ was observed using μ -XRF after salt stress. Increased Cl^- distribution obtained from μ -XRF analysis is an indirect measure of Na^+ as Cl^- is likely associated with Na^+ as a counterion (Figures S8–9). The prevailing higher concentrations of these elements after salt stress are likely to be affecting metabolism and further impacting the observed levels of lipids and other metabolites in a germinating seed. It is worth noting that elemental composition in embryo region of both genotypes change after salt stress. It was observed that the embryo size in both conditions of two genotypes were increased under salt treatment. This is potentially due to role of osmoregulators that exhibit more abscisic acid (ABA) under salt stress. And increase in production of ABA induces the accumulation of more elements to be retained in those regions post-salt treatment (Sun et al., 2015).

Lipids are one of the major components stored exclusively in seed embryos and aleurone layer. They are vital and abundant cellular constituents responsible for cell membrane functionality as well as acting as an energy store to allow metabolism to continue during abiotic stress. The content of lipids in most cereals is relatively low (about 3%) compared to starch and protein. However, their contribution toward the nutritional value, as well as storage stability, of cereal-based food is important (Liu, 2011). Several studies have reported on the content and composition of lipids in whole grains (Price and Parsons, 1975; Welch, 1975; Zhou et al., 1998; Osman et al., 2000; Mehmood et al., 2008) without investigating spatial distributions. In this study, we investigated the effect of salt on the germination of two barley genotypes and on lipid composition and distribution using MALDI-MSI. High quality images taken at a 75 μm raster step size allowed observation of the spatial distribution of a high number of lipids and metabolites in barley seeds. The choice of matrix (DHB) also allowed tentative identification, and profiling, curation and annotation of a large number of lipids. This allowed observation of major lipid species showing differential spatial distributions across seed sections at all time points under control and saline conditions. Despite the higher proportion of lipid species in the embryo [20% lipids (Price and Parsons, 1979; Newman and Newman, 2008)], the largest number of lipids were detected in the endosperm region and not in the embryo. This could be due to the tissue type, as on-tissue extraction of ions is highly influenced by the properties of the tissue. For example, ion abundances can influence the adduct formation, and highly abundant, co-localized lipids or lipids bound to tissue structures can lead to ion suppression (Gorzolka et al., 2016).

It could also be due to the thin tissue content of the embryo at these developmental stages which is highly hydrated (less dense) whereas the endosperm is still very dense.

Lipids play vital roles in providing structural integrity to cell membranes and are involved in signal transduction and membrane trafficking (Krauß and Haucke, 2007). In seeds, lipids serve as a reserve energy source and act as emulsifiers of fat substrates for lipases (Ory et al., 1967). Various environmental stresses including salinity have been shown to be responsible for changes in lipid metabolism and composition (Wang, 2002; Meijer and Munnik, 2003; Mansour, 2013; Natera et al., 2016; Sarabia et al., 2018; Yu et al., 2018). Lipids are rich in olefinic bonds which are susceptible to oxidative attack (Singh and Sinha, 2002; Parida and Das, 2005) from reactive oxygen species (ROSs) present in membranes during salinity stress (Esfandiari et al., 2007). The numbers of annotated species from the GP and ST classes decrease with increasing salinity in both genotypes. This is in line with findings by Wu et al. (2005) where analysis of changes in lipid composition following salt stress was studied in *Spartina patens*, a member of the Poaceae family. The decrease in the number of lipid species in Keel could also be due to peroxidation of lipids as mentioned by Pérez-López et al. (2009).

Among the detected lipid classes in this study, phospholipids were major species of lipids that were tentatively annotated in both genotypes. This was indirectly supported by μ -XRF data where the distribution of P was found across the entire seed of both genotypes (Figures S8–9), correlating with regions where higher numbers of phospholipids were also observed using MSI. Presence of P in the aleurone layer from μ -XRF data also correlates with MALDI-MSI data where higher number of lipids were observed over time. Phospholipid metabolism is associated with a large number of biological molecules and plays a crucial role in various signal transduction pathways in higher plants. In response to various biotic and abiotic stresses (Munnik et al., 1998; Pappan and Wang, 1999), phospholipase D catalyzes the hydrolysis of structural phospholipids, PC and other phospholipids, resulting in the formation of PA. Ritchie and Gilroy (1998) reported that PA is an important mediator of abscisic acid (ABA) signal transduction in barley aleurone protoplasts. Levels of PA were observed to increase slightly with increasing salinity in both genotypes. Salinity stress dependent activation of PA could be due to activation of phospholipase D (PLD) that hydrolyses structural lipids such as PC and PE (Meijer et al., 2002). Increased levels of PA are also proposed to facilitate vesicular trafficking and to recruit target proteins to the membrane, influencing their activity (Testerink and Munnik, 2005; Wang, 2005; Roth, 2008).

High salt concentrations in the growing medium usually lead to an increase in the plasma membrane sterol content (López-Pérez et al., 2009). This increase facilitates hydrophobic interactions among the acyl chains, producing greater order in the membrane, and consequently, a more rigid membrane (Silva et al., 2011). In this study, the percentage of sterols decreased under salt stress in Mundah and Keel seeds. A seed usually has a pair of growing tips (apical meristems), that develop into stem (hypocotyl) and root (radicle). Thus, decrease in sterols in seeds of both genotypes could indicate a need for mobilizing lipids into

meristematic tissue, providing nutrients during germination to support seedling growth during the early stages of development. It was also observed that sterol/phospholipid ratio was greatly affected after salt stress in Keel. Similar results were found in halophyte Brassicaceae species where a significant decrease in sterol lipids was observed after salt stress (Chalbi et al., 2015). However, the sterol/phospholipid ratio was unaffected by salt in Mundah which align with Wu et al. (2005) where the changes in lipid composition after salt stress in *Spartina patens* were studied.

Although phosphatidylcholine levels remained unchanged in Mundah seeds after salt treatment and increased in Keel after exposure to salinity, PC lipid species of 36:n was found at all timepoints in both genotypes under control and saline conditions. Similar results were found in the embryonic axis of *Camelina* seeds detected by MALDI-MSI under normal conditions (Horn et al., 2013; Horn and Chapman, 2014). Increased levels of PCs in response to high salinity was also found by Pical et al. (1999) in *Arabidopsis thaliana*. In plants, PCs are generally produced through a mixed cytidyl diphosphate-choline (CDP-Cho) pathways and methylation pathway (Tasseva et al., 2004). Tasseva et al. (2004) also suggested that enhanced synthesis of PCs is due to an accelerated CDP-Cho pathway. The accumulation of betaine in response to salt has been widely recognized in various plants. *In vitro* studies have proved the positive correlation between the accumulation of betaine and the acquisition of tolerance to salt in maize (*Zea mays*) (Rhodes et al., 1989) and barley (Strange, 1993; Kishitani et al., 1994). Hitz et al. (1981) demonstrated the use of phosphocholine for glycinebetaine synthesis in barley while McDonnell and Jones (1988) also proved that the choline required for oxidation of glycinebetaine is produced *via* turnover of PC. In our study, increased levels of PC over time could possibly result in the synthesis of glycinebetaine, helping seeds to cope under saline conditions. However, this particular mechanism needs further investigation.

High levels of Na^+ also cause secondary responses in plants due to increased oxidative stress. Cellular damage from oxidative stress within plant cells (Apel and Hirt, 2004) induces production of reactive oxygen species (ROSs) (Jaleel et al., 2007; Ashraf, 2009). ROSs derived from molecular oxygen can accumulate in the plant cell and cause oxidative damage in cellular components, including proteins and lipids. To prevent the potential cytotoxic effects of ROSs, the stimulation of antioxidant systems can assist in plant protection from oxidative stress (Grace and Logan, 2000; Chutipajit et al., 2009). Polyphenolic compounds such as flavonoids play an important role in stopping the propagation of oxidative chain reactions (Ksouri et al., 2007). Their synthesis is generally stimulated in response to various abiotic stresses including salinity (Zhou et al., 2018) and thus the biosynthesis of such compounds is generally stimulated in salt-exposed plants. Flavonoids are also potential inhibitors of the enzyme lipoxygenase, which converts polyunsaturated fatty acids to oxygen-containing derivatives (Nijveldt et al., 2001). These compounds accumulate in plant tissues protecting them from damaging effects of salt stress and inhibiting lipid peroxidation (Potapovich and Kostyuk, 2003). In our MSI results, a tentative flavonoid (m/z 365.102 [$\text{M}+\text{H}^+$]) was found to be discriminant in Mundah and Keel embryos after exposure to salinity. Analysis

using LC-MS/MS and MALDI-MS/MS confirmed the compound to be a flavonoid, tentatively annotated as Gaiconin F. Flavonoid profiling data from LC-QToF-MS supported our MALDI-MSI data (refer to **Figure S7** and **Table S11**). This is in alignment with Chutipaijit et al. (2009) who described a significant increase in total flavonoid content in salt stressed rice seedlings and Taïbi et al. (2016) where a significant increase in flavonoid content was seen in *Phaseolus vulgaris* under salt stress.

Under normal conditions, the quiescent embryo is able to germinate after imbibition and breaking dormancy (McDonald and Copeland, 2012). However, in salt stress conditions, the radicle has to avoid oxidative damage for successful germination. To detoxify or scavenge the severe effects of stress, the presence of antioxidants in embryos may play a crucial role in protecting the emerging radicle and help in a successful germination. This is also supported by Shirley (1998), who described the accumulation of flavonoids in the embryos of all plant species. Our observed detection of a tentative flavonoid in the embryos of Munda and Keel supports the view of flavonoids playing potential roles in the protection of seeds under salt stress. Keel displayed a significantly delayed flavonoid response relative to Munda as shown in the LC-MS data where highest X-fold increase in Keel was observed at 72 h. This depicts its lesser ability to deal with salt leading to poorer germination efficiency. However, in Munda, flavonoid content was highest at 48 h displaying faster responses to salinity correlating with the higher germination efficiency. These results indicate a possible advantage for the Munda genotype, making it a better germinator under salt stress. Gradual increase in the number of lipids detected over time from MALDI-MSI in Keel compared to Munda under salt conditions also support our findings of Keel showing slow germination.

Analysis of the flavonoid profile of the five other barley genotypes showed varied and contrasting responses under salt stress over time. Vlammingh, which showed a germination efficiency between that of Keel and Munda, also showed an increase in the Gancaonin F flavonoid under salt stress to a level between that of Keel and Munda. These observations point to Gancaonin F providing some advantage to specific genotypes in the response to salt stress through possible antioxidant scavenging of ROS species and protection of the vulnerable embryo during germination. The other four genotypes showed a variety of differing patterns in Gancaonin F profiles. There are a multitude of possible response mechanisms that the germinating seed is able to employ and the varying responses observed are most likely due to the wide range in genetic diversity of the selected barley varieties.

CONCLUSIONS

This study investigated the changes in levels and distribution of metabolites and elements in germinating seeds of Munda and Keel, two barley genotypes that have contrasting germination rates in response to salt stress. To compare and contrast the lipid and metabolite profiles of these varieties during the early stages of seed germination under control and saline conditions,

several orthogonal approaches were combined using MALDI-MSI, elemental composition analysis using ICP-MS, spatial distribution analysis using μ -XRF spectrometry and confirmation of compounds using LC-QToF-MS. Use of discriminative MALDI-MSI analysis as an exploratory and qualitative technique allowed visualization of larger lipid differences within different seeds structures as well as detection of a tentative flavonoid that showed discriminate behaviour between the two genotypes at different time points. Further analysis across genotypes to determine the roles of flavonoids in barley germination under salt stress conditions is required. However, these initial results indicate flavonoid as a strong candidate metabolite biomarker for detecting salinity tolerant varieties.

DATA AVAILABILITY

This manuscript contains previously unpublished data. The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

SG carried out the experiments, data analysis and interpretation of the results. SG wrote the manuscript with support from BB, SN, PS, CH, UR. LCMS was conducted and analyzed by SG and TR. ICP-MS was conducted and analyzed by SG and DC. μ -XRF imaging was conducted by BB. Both BB and UR supervised the project. All authors provided critical feedback, helped shape the research and authorized the final manuscript.

FUNDING

This project and UR were funded through an Australian Research Council (ARC) Future Fellowship program (Grant Number FT130100326).

ACKNOWLEDGMENTS

The authors would like to thank The University of Adelaide for providing the barley seeds. Jens Burgmann (Bruker Pty Ltd, Mining and Geo Sciences, Darra, Qld) for conducting μ -XRF imaging of seed sections. LC-MS and MALDI-MSI were carried out at Metabolomics Australia which is supported by funds from the Australian Government's National Collaborative Research Infrastructure Scheme (NCRIS) administered through Bioplatforms Australia (BPA) Ltd.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.01139/full#supplementary-material>

REFERENCES

- Achakzai, A. K. K. (2014). Imbibition, germination and lipid mobilization response by sunflower subjected to salt stress. *Pak J. Bot.* 46, 1995–2002.
- Ahmad, J. (1992). The effect of sodium chloride on the physiology of cotyledons and mobilization of reserve food in *Cicer arietinum*. *Pak. J. Bot.* 24, 40–48.
- Alencar, N. L., Gadelha, C. G., Gallão, M. I., Dolder, M. A., Prisco, J. T., and Gomes-Filho, E. (2015). Ultrastructural and biochemical changes induced by salt stress in *Jatropha curcas* seeds during germination and seedling development. *Funct. Plant Biol.* 42, 865–874. doi: 10.1071/FP15019
- Apel K. and Hirt H. (2004). Reactive oxygen species: metabolism, oxidative stress, and signal transduction. *Annu. Rev. Plant Biol.* 55, 373–399. doi: 10.1146/annurev.arplant.55.031903.141701
- Ashraf, M. (2009). Biotechnological approach of improving plant salt tolerance using antioxidants as markers. *Biotechnol. Adv.* 27, 84–93. doi: 10.1016/j.biotechadv.2008.09.003
- Ashraf, M., Zafar, R., and Ashraf, M. Y. (2003). Time-course changes in the inorganic and organic components of germinating sunflower achenes under salt (NaCl) stress. *Flora (Jena) [DE-600]*. 198, 26–36. doi: 10.1078/0367-2530-00073
- Bartwal, A., Mall, R., Lohani, P., Guru, S., and Arora, S. J. (2013). Role of secondary metabolites and brassinosteroids in plant defense against environmental stresses. *J. Plant Growth Regul.* 32, 216–232. doi: 10.1007/s00344-012-9272-x
- Begum, F., Karmoker, J., Fattah, Q., and Maniruzzaman, A. (1992). The effect of salinity on germination and its correlation with K⁺, Na⁺, Cl⁻ accumulation in germinating seeds of *Triticum aestivum* L. cv. Akbar. *Plant Cell Physiol.* 33, 1009–1014. doi: 10.1093/oxfordjournals.pcp.a078324
- Bewley, J. D. (1997). Seed germination and dormancy. *Plant Cell* 9, 1055. doi: 10.1105/tpc.9.7.1055
- Bewley, J. D. (2001). Seed germination and reserve mobilization. In eLS, (Ed.). doi: 10.1038/npg.els.0002047
- Bewley, J. D., and Black, M. (1994). *Seeds*. New York: Springer Science & Business Media, 1–33. doi: 10.1007/978-1-4899-1002-8_1
- Bliss, R., Platt-Aloia, K., and Thomson, W. (1986). Osmotic sensitivity in relation to salt sensitivity in germinating barley seeds. *Plant Cell Environ.* 9, 721–725. doi: 10.1111/j.1365-3040.1986.tb02104.x
- Boughton, B. A., Thinagaran, D., Sarabia, D., Bacic, A., and Roessner, U. (2016). Mass spectrometry imaging for plant biology: a review. *Phytochem. Rev.* 15, 445–488. doi: 10.1007/s11101-015-9440-2
- Burrell, M., Earnshaw, C., and Clench, M. (2006). Imaging matrix assisted laser desorption ionization mass spectrometry: a technique to map plant metabolites within tissues at high spatial resolution. *J. Exp. Bot.* 58, 757–763. doi: 10.1093/jxb/erl139
- Callahan, D. L., Hare, D. J., Bishop, D. P., Doble, P. A., and Roessner, U. (2016). Elemental imaging of leaves from the metal hyperaccumulating plant *Noccaea caerulescens* shows different spatial distribution of Ni, Zn and Cd. *RSC Adv.* 6, 2337–2344. doi: 10.1039/C5RA23953B
- Calvano, C. D., Cataldi, T. R., Kögel, J. F., Monopoli, A., Palmisano, E., and Sundermeyer, J. (2017). Structural characterization of neutral saccharides by negative ion MALDI mass spectrometry using a superbasic proton sponge as deprotonating matrix. *J. Am. Soc. Mass Spectrom.* 28, 1666–1675. doi: 10.1007/s13361-017-1679-y
- Cao, D., Lutz, A., Hill, C. B., Callahan, D. L., and Roessner, U. (2017). A quantitative profiling method of phytohormones and other metabolites applied to barley roots subjected to salinity stress. *Front. Plant Sci.* 7, 2070. doi: 10.3389/fpls.2016.02070
- Cevallos-Cevallos, J. M., Reyes-De-Corcuera, J. I., Etxeberria, E., Danyluk, M. D., and Rodrick, G. E. (2009). Metabolomic analysis in food science: a review. *Trends Food Sci. Technol.* 20, 557–566. doi: 10.1016/j.tifs.2009.07.002
- Chalbi, N., Martínez-Ballesta, M. C., Youssef, N. B., and Carvajal, M. (2015). Intrinsic stability of Brassicaceae plasma membrane in relation to changes in proteins and lipids as a response to salinity. *J. Plant Physiol.* 175, 148–156. doi: 10.1016/j.jplph.2014.12.003
- Chutipajit, S., Cha-Um, S., and Sompornpailin, K. (2009). Differential accumulations of proline and flavonoids in indica rice varieties against salinity. *Pak. J. Bot.* 41, 2497–2506.
- Cornett, D. S., Reyzer, M. L., Chaurand, P., and Caprioli, R. M. (2007). MALDI imaging mass spectrometry: molecular snapshots of biochemical systems. *Nat. Methods* 4, 828–833. doi: 10.1038/nmeth1094
- Esfandiari, E., Shekari, F., Shekari, F., and Esfandiari, M. (2007). The effect of salt stress on antioxidant enzymes activity and lipid peroxidation on the wheat seedling. *Not. Bot. Horti Agrobi.* 35, 48. doi: 10.15835/nbha.35.1.251
- Fahy, E., Sud, M., Cotter, D., and Subramaniam, S. (2007). LIPID MAPS online tools for lipid research. *Nucleic Acids Res.* 35, W606–W612. doi: 10.1093/nar/gkm324
- Fait, A., Angelovici, R., Less, H., Ohad, I., Urbanczyk-Wochniak, E., Fernie, A. R., et al. (2006). Arabidopsis seed development and germination is associated with temporally distinct metabolic switches. *Plant Physiol.* 142, 839–854. doi: 10.1104/pp.106.086694
- Fiehn, O. (2002). Metabolomics—the link between genotypes and phenotypes. *Plant Mol. Biol.* 48, 155–171. doi: 10.1023/A:1013713905833
- Flowers, T., Troke, P., and Yeo, A. (1977). The mechanism of salt tolerance in halophytes. *Annu. Rev. Plant Physiol.* 28, 89–121. doi: 10.1146/annurev.pp.28.060177.000513
- Fulcher, R. (1982). Fluorescence microscopy of cereals. *Food Microstruct* 1, 7.
- Gorzolka, K., Bednarz, H., and Niehaus, K. (2014). Detection and localization of novel hordatine-like compounds and glycosylated derivatives of hordatines by imaging mass spectrometry of barley seeds. *Planta* 239, 1321–1335. doi: 10.1007/s00425-014-2061-y
- Gorzolka, K., Kölling, J., Nattkemper, T. W., and Niehaus, K. (2016). Spatio-Temporal Metabolite Profiling of the Barley Germination Process by MALDI MS Imaging. *PLoS One* 11, e0150208. doi: 10.1371/journal.pone.0150208
- Grace, S. C., and Logan, B. A. (2000). Energy dissipation and radical scavenging by the plant phenylpropanoid pathway. *Philos. Trans. R. Soc. Lond., B, Biol.* 355, 1499–1510. doi: 10.1098/rstb.2000.0710
- Gry, J., Black, L., Eriksen, F. D., Pilegaard, K., Plumb, J., Rhodes, M., et al. (2007). EuroFIR-BASIS — a combined composition and biological activity database for bioactive compounds in plant-based foods. *Trends Food Sci. Technol.* 18, 434–444. doi: 10.1016/j.tifs.2007.05.008
- Hampson, C. R., and Simpson, G. (1990). Effects of temperature, salt, and osmotic potential on early growth of wheat (*Triticum aestivum*). I. Germination. *Can. J. Bot.* 68, 524–528. doi: 10.1139/b90-072
- Hansen, C. C., Sørensen, M., Veiga, T. A., Zibrandtsen, J. F., Heskes, A. M., Olsen, C. E., et al. (2018). Reconfigured cyanogenic glucoside biosynthesis in *Eucalyptus cladocalyx* involves a cytochrome P450 CYP706C55. *Plant Physiol.* 178, 1081–1095. doi: 10.1104/pp.18.00998
- Hasanuzzaman, M., Nahar, K., and Fujita, M. (2013). “Plant response to salt stress and role of exogenous protectants to mitigate salt-induced damages,” in *Ecophysiology and Responses of plants under salt stress* New York: Springer, 25–87. doi: 10.1007/978-1-4614-4747-4_2
- Hasegawa, P. M. (2013). Sodium (Na⁺) homeostasis and salt tolerance of plants. *Environ. Exp. Bot.* 92, 19–31. doi: 10.1016/j.envexpbot.2013.03.001
- Hasegawa, P. M., Bressan, R. A., Zhu, J.-K., and Bohnert, H. J. (2000). Plant cellular and molecular responses to high salinity. *Annu. Rev. Plant Biol.* 51, 463–499. doi: 10.1146/annurev.arplant.51.1.463
- Heskes, A. M., Sundram, T. C., Boughton, B. A., Jensen, N. B., Hansen, N. L., Crocoll, C., et al. (2018). Biosynthesis of bioactive diterpenoids in the medicinal plant *Vitex agnus-castus*. *Plant J.* 93, 943–958. doi: 10.1111/tj.13822
- Hitz, W. D., Rhodes, D., and Hanson, A. D. (1981). Radiotracer evidence implicating phosphoryl and phosphatidyl bases as intermediates in betaine synthesis by water-stressed barley leaves. *Plant Physiol.* 68, 814–822. doi: 10.1104/pp.68.4.814
- Horn, P. J., and Chapman, K. D. (2014). Lipidomics in situ: insights into plant lipid metabolism from high resolution spatial maps of metabolites. *Prog. Lipid. Res.* 54, 32–52. doi: 10.1016/j.plipres.2014.01.003
- Horn, P. J., Silva, J. E., Anderson, D., Fuchs, J., Borisjuk, L., Nazarenus, T. J., et al. (2013). Imaging heterogeneity of membrane and storage lipids in transgenic *C. melinensis* seeds with altered fatty acid profiles. *Plant J.* 76, 138–150. doi: 10.1111/tj.12278
- Howell, K. A., Narsai, R., Carroll, A., Ivanova, A., Lohse, M., Usadel, B., et al. (2009). Mapping metabolic and transcript temporal switches during germination in rice highlights specific transcription factors and the role of RNA instability in the germination process. *Plant Physiol.* 149, 961–980. doi: 10.1104/pp.108.129874
- Huang, C. X., and Steveninck, R. F. (1988). Effect of moderate salinity on patterns of potassium, sodium and chloride accumulation in cells near the root tip of barley: role of differentiating metaxylem vessels. *Physiol. Plant* 73, 525–533. doi: 10.1111/j.1399-3054.1988.tb05436.x

- Jakobsen Neilson, E. H., Marques Dos Santos, B., Sørensen Zibrandtsen, J. F., Disan, G., Sørensen, M., Cozzi, F., et al. (2019). Quantification and localization of formylated phloroglucinol compounds (FPCs) in Eucalyptus species. *Front. Plant Sci.* 10, 186. doi: 10.3389/fpls.2019.00186
- Jaleel, C. A., Gopi, R., Sankar, B., Manivannan, P., Kishorekumar, A., Sridharan, R., et al. (2007). Studies on germination, seedling vigour, lipid peroxidation and proline metabolism in *Catharanthus roseus* seedlings under salt stress. *S. Afr. J. Bot.* 73, 190–195. doi: 10.1016/j.sajb.2006.11.001
- Kamboj, A., Ziemann, M., and Bhave, M. (2015). Identification of salt-tolerant barley varieties by a consolidated physiological and molecular approach. *Acta Physiol. Plant.* 37, 1716. doi: 10.1007/s11738-014-1716-4
- Karl, T. R., Melillo, J., and Peterson, T. (2009). *Global climate change impacts in the United States*. The United States of America: Cambridge University Press.
- Kawamoto, T. (2003). Use of a new adhesive film for the preparation of multipurpose fresh-frozen sections from hard tissues, whole animals, insects and plants. *Arch. Histol. Cytol.* 66, 123–143. doi: 10.1679/aohc.66.123
- Khan, M. A., and Rizvi, Y. (1994). Effect of salinity, temperature, and growth regulators on the germination and early seedling growth of *Atriplex griffithii* var. stocksii. *Can. J. Bot.* 72, 475–479. doi: 10.1139/b94-063
- Kishitani, S., Watanabe, K., Yasuda, S., Arakawa, K., and Takabe, T. (1994). Accumulation of glycinebetaine during cold acclimation and freezing tolerance in leaves of winter and spring barley plants. *Plant Cell Environ.* 17, 89–95. doi: 10.1111/j.1365-3040.1994.tb00269.x
- Krauß, M., and Haucke, V. (2007). Phosphoinositide-metabolizing enzymes at the interface between membrane traffic and cell signalling. *EMBO Rep.* 8, 241–246. doi: 10.1038/sj.embor.7400919
- Ksouri, R., Megdiche, W., Debez, A., Falleh, H., Grignon, C., and Abdelly, C. (2007). Salinity effects on polyphenol content and antioxidant activities in leaves of the halophyte *Cakile maritima*. *Plant Physiol. Biochem.* 45, 244–249. doi: 10.1016/j.plaphy.2007.02.001
- Lee, D. Y., Bowen, B. P., and Northen, T. R. (2010). Mass spectrometry-based metabolomics, analysis of metabolite-protein interactions, and imaging. *Bio. Techniques* 49, 557–565. doi: 10.2144/000113451
- Lee, Y. J., Perdian, D. C., Song, Z., Yeung, E. S., and Nikolau, B. J. (2012). Use of mass spectrometry for imaging metabolites in plants. *Plant J.* 70, 81–95. doi: 10.1111/j.1365-313X.2012.04899.x
- Li, Y., Shrestha, B., and Vertes, A. (2008). Atmospheric pressure infrared MALDI imaging mass spectrometry for plant metabolomics. *Anal. Chem.* 80, 407–420. doi: 10.1021/ac701703f
- Liebisch, G., Vizcaino, J. A., Köfeler, H., Trötz Müller, M., Griffiths, W. J., Schmitz, G., et al. (2013). Shorthand notation for lipid structures derived from mass spectrometry. *J. Lipid Res.* 54, 1523–1530. doi: 10.1194/jlr.M033506
- Liu, K. (2011). Comparison of lipid content and fatty acid composition and their distribution within seeds of 5 small grain species. *J. Food Sci.* 76, C334–342. doi: 10.1111/j.1750-3841.2010.02038.x
- Long, N., Jefferies, S., Warner, P., Karakousis, A., Kretschmer, J., Hunt, C., et al. (2013). Mapping and QTL analysis of the barley population Mundahx Keel. *Aust. J. Agric. Res.* 54, 1163–1171. doi: 10.1071/AR02206
- López-Pérez, L., Del Carmen Martínez-Ballesta, M., Maurel, C., and Carvajal, M. (2009). Changes in plasma membrane lipids, aquaporins and proton pump of broccoli roots, as an adaptation mechanism to salinity. *Phytochemistry* 70, 492–500. doi: 10.1016/j.phytochem.2009.01.014
- Mansour, M. (2013). Plasma membrane permeability as an indicator of salt tolerance in plants. *Biol. Plant.* 57, 1–10. doi: 10.1007/s10535-012-0144-9
- Marques, E. C., De Freitas, P. A. F., Alencar, N. L. M., Prisco, J. T., and Gomes-Filho, E. (2013). Increased Na⁺ and Cl[−] accumulation induced by NaCl salinity inhibits cotyledonary reserve mobilization and alters the source-sink relationship in establishing dwarf cashew seedlings. *Acta Physiol. Plant.* 35, 2171–2182. doi: 10.1007/s11738-013-1254-5
- McDonald, M., and Copeland, L. O. (2012). *seed production: principles and practices*. New York: Springer Science & Business Media.
- McDonnell, E., and Jones, R. W. (1988). Glycinebetaine biosynthesis and accumulation in unstressed and salt-stressed wheat. *J. Exp. Bot.* 39, 421–430. doi: 10.1093/jxb/39.4.421
- Mehmood, S., Orhan, I., Ahsan, Z., Aslan, S., and Gulfray, M. (2008). Fatty acid composition of seed oil of different Sorghum bicolor varieties. *Food Chem.* 109, 855–859. doi: 10.1016/j.foodchem.2008.01.014
- Meijer, H. J., and Munnik, T. (2003). Phospholipid-based signaling in plants. *Annu. Rev. Plant Biol.* 54, 265–306. doi: 10.1146/annurev.arplant.54.031902.134748
- Meijer, H. J., Ter Riet, B., Van Himbergen, J. A., Musgrave, A., and Munnik, T. (2002). KCl activates phospholipase D at two different concentration ranges: distinguishing between hyperosmotic stress and membrane depolarization. *Plant J.* 31, 51–60. doi: 10.1046/j.1365-313X.2002.01336.x
- Mullen, A. K., Clench, M. R., Crosland, S., and Sharples, K. R. (2005). Determination of agrochemical compounds in soya plants by imaging matrix-assisted laser desorption/ionisation mass spectrometry. *Rapid Commun. Mass Spectrom.* 19, 2507–2516. doi: 10.1002/rcm.2078
- Munnik, T., Irvine, R., and Musgrave, A. (1998). Phospholipid signalling in plants. *Biochim. Biophys. Acta* 1389, 222–272. doi: 10.1016/S0005-2760(97)00158-6
- Munns, R., and Tester, M. (2008). Mechanisms of salinity tolerance. *Annu Rev. Plant Biol.* 59, 651–681. doi: 10.1146/annurev.arplant.59.032607.092911
- Nakabayashi, K., Okamoto, M., Koshiba, T., Kamiya, Y., and Nambara, E. (2005). Genome-wide profiling of stored mRNA in *Arabidopsis thaliana* seed germination: epigenetic and genetic regulation of transcription in seed. *Plant J.* 41, 697–709. doi: 10.1111/j.1365-313X.2005.02337.x
- Natera, S. H., Hill, C. B., Rupasinghe, T. W., and Roessner, U. (2016). Salt-stress induced alterations in the root lipidome of two barley genotypes with contrasting responses to salinity. *Funct. Plant Biol.* 43, 207–219. doi: 10.1071/FP15253
- Neumann, P. (1997). Salinity resistance and plant growth revisited. *Plant Cell Environ.* 20, 1193–1198. doi: 10.1046/j.1365-3040.1997.d01-139.x
- Newman, R. K., and Newman, C. W. (2008). *Barley for food and health: Science, technology, and products*. United States: John Wiley & Sons. doi: 10.1002/9780470369333
- Nijveldt, R. J., Van Nood, E., Van Hoorn, D. E., Boelens, P. G., Van Norren, K., and Van Leeuwen, P. A. (2001). Flavonoids: a review of probable mechanisms of action and potential applications. *Am. J. Clin. Nutr.* 74, 418–425. doi: 10.1093/ajcn/74.4.418
- Okazaki, Y., and Saito, K. (2014). Roles of lipids as signaling molecules and mitigators during stress response in plants. *Plant J.* 79, 584–596. doi: 10.1111/tpj.12556
- Ory, R. L., St. Angelo, A. J., Degruy, I. V., and Altschul, A. M. (1967). Castor bean lipase: Role of the lipid cofactor. *Can. J. Biochem.* 45, 1445–1450. doi: 10.1139/o67-170
- Osman, R., El-Gelil, F. A., El-Noamany, H., and Dawood, M. (2000). Oil content and fatty acid composition of some varieties of barley and sorghum grains. *GRASAS ACEITES* 51, 157–162. doi: 10.3989/gya.2000.v51.i3.472
- Pappan, K., and Wang, X. (1999). Molecular and biochemical properties and physiological roles of plant phospholipase D. *Biochim. Biophys. Acta* 1439, 151–166. doi: 10.1016/S1388-1981(99)00091-8
- Pardo, J. M., Cubero, B., Leidi, E. O., and Quintero, F. J. (2006). Alkali cation exchangers: roles in cellular homeostasis and stress tolerance. *J. Exp. Bot.* 57, 1181–1199. doi: 10.1093/jxb/erj114
- Parida, A. K., and Das, A. B. (2005). Salt tolerance and salinity effects on plants: a review. *Ecotoxicol. Environ. Saf.* 60, 324–349. doi: 10.1016/j.ecoenv.2004.06.010
- Pérez-López, U., Robredo, A., Lacuesta, M., Sgherri, C., Muñoz-Rueda, A., Navari-Izzo, F., et al. (2009). The oxidative stress caused by salinity in two barley cultivars is mitigated by elevated CO₂. *Physiol. Plant* 135, 29–42. doi: 10.1111/j.1399-3054.2008.01174.x
- Pical, C., Westergren, T., Dove, S. K., Larsson, C., and Sommarin, M. (1999). Salinity and hyperosmotic stress induce rapid increases in phosphatidylinositol 4, 5-bisphosphate, diacylglycerol pyrophosphate, and phosphatidylcholine in *Arabidopsis thaliana* cells. *J. Biol. Chem.* 274, 38232–38240. doi: 10.1074/jbc.274.53.38232
- Potapovich, A., and Kostyuk, V. (2003). Comparative study of antioxidant properties and cytoprotective activity of flavonoids. *Biochem. Mosc.* 68, 514–519. doi: 10.1023/A:1023947424341
- Price, P. B., and Parsons, J. (1979). Distribution of lipids in embryonic axis, bran-endosperm, and hull fractions of hullless barley and hullless oat grain. *J. Agric. Food Chem.* 27, 813–815. doi: 10.1021/jf60224a030
- Price, P. B., and Parsons, J. G. (1975). Lipids of seven cereal grains. *J. Am. Oil Chem. Soc.* 52, 490–493. doi: 10.1007/BF02640738
- Rauscher, S., Marquardt, C., Balluff, B., Deininger, S. R.-O., Albers, C., Belau, E., et al. (2010). Classification of HER2 receptor status in breast cancer tissues by MALDI imaging mass spectrometry. *J. Proteome Res.* 9, 1854–1863. doi: 10.1021/pr901008d
- Reyes, J. L., and Chua, N. H. (2007). ABA induction of miR159 controls transcript levels of two MYB factors during *Arabidopsis* seed germination. *Plant J.* 49, 592–606. doi: 10.1111/j.1365-313X.2006.02980.x

- Rhodes, D., Rich, P. J., Brunk, D. G., Ju, G. C., Rhodes, J. C., Pauly, M. H., et al. (1989). Development of two isogenic sweet corn hybrids differing for glycinebetaine content. *Plant Physiol.* 91, 1112–1121. doi: 10.1104/pp.91.3.1112
- Ritchie, S., and Gilroy, S. (1998). Abscissic acid signal transduction in the barley aleurone is mediated by phospholipase D activity. *Proc. Natl. Acad. Sci.* 95, 2697–2702. doi: 10.1073/pnas.95.5.2697
- Robinson, S., Warburton, K., Seymour, M., Clench, M., and Thomas-Oates, J. (2007). Localization of water-soluble carbohydrates in wheat stems using imaging matrix-assisted laser desorption/ionization mass spectrometry. *New Phytol.* 173, 438–444. doi: 10.1111/j.1469-8137.2006.01934.x
- Roth, M. G. (2008). Molecular mechanisms of PLD function in membrane traffic. *Traffic* 9, 1233–1239. doi: 10.1111/j.1600-0854.2008.00742.x
- Routaboul, J. M., Kerhoas, L., Debeaujon, I., Pourcel, L., Caboche, M., Einhorn, J., et al. (2006). Flavonoid diversity and biosynthesis in seed of *Arabidopsis thaliana*. *Planta* 224, 96–107. doi: 10.1007/s00425-005-0197-5
- Sarabia, L. D., Boughton, B. A., Rupasinghe, T., Van De Meene, A. M., Callahan, D. L., Hill, C. B., et al. (2018). High-mass-resolution MALDI mass spectrometry imaging reveals detailed spatial distribution of metabolites and lipids in roots of barley seedlings in response to salinity stress. *Metabolomics* 14, 63. doi: 10.1007/s11306-018-1359-3
- Shabala, S., and Cuin, T. A. (2008). Potassium transport and plant salt tolerance. *Physiol. Plant* 133, 651–669. doi: 10.1111/j.1399-3054.2007.01008.x
- Shelden, M. C., Dias, D. A., Jayasinghe, N. S., Bacic, A., and Roessner, U. (2016). Root spatial metabolite profiling of two genotypes of barley (*Hordeum vulgare* L.) reveals differences in response to short-term salt stress. *J. Exp. Bot.* 67, 3731–3745. doi: 10.1093/jxb/erw059
- Shelden, M. C., and Roessner, U. (2013). Advances in functional genomics for investigating salinity stress tolerance mechanisms in cereals. *Front. Plant Sci.* 4, 123. doi: 10.3389/fpls.2013.00123
- Shelden, M. C., Roessner, U., Sharp, R. E., Tester, M., and Bacic, A. (2013). Genetic variation in the root growth response of barley genotypes to salinity stress. *Funct. Plant Biol.* 40, 516–530. doi: 10.1071/FP12290
- Shirley, B. W. (1998). Flavonoids in seeds and grains: physiological function, agronomic importance and the genetics of biosynthesis. *Seed Sci. Res.* 8, 415–422. doi: 10.1017/S0960258500004372
- Shroff, R., Vergara, F., Muck, A., Svatoš, A., and Gershenzon, J. (2008). Nonuniform distribution of glucosinolates in *Arabidopsis thaliana* leaves has important consequences for plant defense. *Proc. Natl. Acad. Sci.* 105, 6196–6201. doi: 10.1073/pnas.0711730105
- Silva, C., Aranda, F. J., Ortiz, A., Martínez, V., Carvajal, M., and Teruel, J. A. (2011). Molecular aspects of the interaction between plants sterols and DPPC bilayers: an experimental and theoretical approach. *J. Colloid Interface Sci.* 358, 192–201. doi: 10.1016/j.jcis.2011.02.048
- Singh, S. C., and Sinha, R. P. (2002). Role of lipids and fatty acids in stress tolerance in cyanobacteria. *Acta Protozool.* 41, 297–308.
- Smith, C. A., O'maille, G., Want, E. J., Qin, C., Trauger, S. A., Brandon, T. R., et al. (2005). METLIN: a metabolite mass spectral database. *Ther. Drug Monit.* 27, 747–751. doi: 10.1097/01.fid.0000179845.53213.39
- Strange, K. (1993). *Cellular and molecular physiology of cell volume regulation*. United States: CRC Press.
- Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., et al. (2007). Proposed minimum reporting standards for chemical analysis. *Metabolomics* 3, 211–221. doi: 10.1007/s11306-007-0082-2
- Sumner, L. W., Mendes, P., and Dixon, R. A. (2003). Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry* 62, 817–836. doi: 10.1016/S0031-9422(02)00708-2
- Sun, Y., Kong, X., Li, C., Liu, Y., and Ding, Z. J. P. O. (2015). Potassium retention under salt stress is associated with natural variation in salinity tolerance among *Arabidopsis* accessions. *PLoS One* 10, e0124032. doi: 10.1371/journal.pone.0124032
- Taibi, K., Taibi, F., Abderrahim, L. A., Ennajah, A., Belkhdja, M., and Mulet, J. M. (2016). Effect of salt stress on growth, chlorophyll content, lipid peroxidation and antioxidant defence systems in *Phaseolus vulgaris* L. *S. Afr. J. Bot.* 105, 306–312. doi: 10.1016/j.sajb.2016.03.011
- Tasseva, G., Richard, L., and Zachowski, A. (2004). Regulation of phosphatidylcholine biosynthesis under salt stress involves choline kinases in *Arabidopsis thaliana*. *FEBS Lett.* 566, 115–120. doi: 10.1016/j.febslet.2004.04.015
- Tavakkoli, E., Fatehi, F., Rengasamy, P., and McDonald, G. K. (2012). A comparison of hydroponic and soil-based screening methods to identify salt tolerance in the field in barley. *J. Exp. Bot.* 63, 3853–3867. doi: 10.1093/jxb/ers085
- Testerink, C., and Munnik, T. (2005). Phosphatidic acid: a multifunctional stress signaling lipid in plants. *Trends Plant Sci.* 10, 368–375. doi: 10.1016/j.tplants.2005.06.002
- Volkov, V., and Amtmann, A. (2006). *Thellungiella halophila*, a salt-tolerant relative of *Arabidopsis thaliana*, has specific root ion-channel features supporting K⁺/Na⁺ homeostasis under salinity stress. *Plant J.* 48, 342–353. doi: 10.1111/j.1365-313X.2006.02876.x
- Wang, X. (1999). The role of phospholipase D in signaling cascades. *Plant Physiol.* 120, 645–652. doi: 10.1104/pp.120.3.645
- Wang, X. (2002). Phospholipase D in hormonal and stress signaling. *Curr. Opin. Plant Biol.* 5, 408–414. doi: 10.1016/S1369-5266(02)00283-2
- Wang, X. (2005). Regulatory functions of phospholipase D and phosphatidic acid in plant growth, development, and stress responses. *Plant Physiol.* 139, 566–573. doi: 10.1104/pp.105.068809
- Weckwerth, W. (2003). Metabolomics in systems biology. *Annu. Rev. Plant Biol.* 54, 669–689. doi: 10.1146/annurev.arplant.54.031902.135014
- Welch, R. W. (1975). Fatty acid composition of grain from winter and spring sown oats, barley and wheat. *J. Sci. Food Agric.* 26, 429–435. doi: 10.1002/jsfa.2740260408
- Wu, J., Seliskar, D. M., and Gallagher, J. L. (2005). The response of plasma membrane lipid composition in callus of the halophyte *Spartina patens* (Poaceae) to salinity stress. *Am. J. Bot.* 92, 852–858. doi: 10.3732/ajb.92.5.852
- Yang, P., Li, X., Wang, X., Chen, H., Chen, F., and Shen, S. (2007). Proteomic analysis of rice (*Oryza sativa*) seeds during germination. *Proteomics* 7, 3358–3368. doi: 10.1002/pmic.200700207
- Ye, H., Gemperline, E., Venkateshwaran, M., Chen, R., Delaux, P. M., Howes-Podoll, M., et al. (2013). MALDI mass spectrometry-assisted molecular imaging of metabolites during nitrogen fixation in the *Medicago truncatula*–*Sinorhizobium meliloti* symbiosis. *Plant J.* 75, 130–145. doi: 10.1111/tpj.12191
- Yu, D., Rupasinghe, T. W., Boughton, B. A., Natera, S. H., Hill, C. B., Tarazona, P., et al. (2018). A high-resolution HPLC-QqTOF platform using parallel reaction monitoring for in-depth lipid discovery and rapid profiling. *Anal. Chim. Acta.* 1026, 87–100. doi: 10.1016/j.aca.2018.03.062
- Zaima, N., Goto-Inoue, N., Hayasaka, T., and Setou, M. (2010a). Application of imaging mass spectrometry for the analysis of *Oryza sativa* rice. *Rapid Commun. Mass Spectrom.* 24, 2723–2729. doi: 10.1002/rcm.4693
- Zaima, N., Hayasaka, T., Goto-Inoue, N., and Setou, M. (2009). Imaging of metabolites by MALDI mass spectrometry. *J. Oleo Sci.* 58, 415–419. doi: 10.5650/jos.58.415
- Zaima, N., Hayasaka, T., Goto-Inoue, N., and Setou, M. (2010b). Matrix-assisted laser desorption/ionization imaging mass spectrometry. *Int. J. Mol. Sci.* 11, 5040–5055. doi: 10.3390/ijms11125040
- Zhang, H., Irving, L. J., McGill, C., Matthew, C., Zhou, D., and Kemp, P. (2010). The effects of salinity and osmotic stress on barley germination rate: sodium as an osmotic regulator. *Ann. Bot.* 106, 1027–1035. doi: 10.1093/aob/mcq204
- Zhou, M., Holmes, M. G., Robards, K., and Helliwell, S. (1998). Fatty acid composition of lipids of Australian oats. *J. Cereal Sci.* 28, 311–319. doi: 10.1016/S0733-5210(98)90011-X
- Zhou, Y., Tang, N., Huang, L., Zhao, Y., Tang, X., and Wang, K. (2018). Effects of Salt Stress on Plant Growth, Antioxidant Capacity, Glandular Trichome Density, and Volatile Exudates of *Schizonepeta tenuifolia* Briq. *Int. J. Mol. Sci.* 19, 252. doi: 10.3390/ijms19010252
- Zhu, J.-K. (2003). Regulation of ion homeostasis under salt stress. *Curr. Opin. Plant Biol.* 6, 441–445. doi: 10.1016/S1369-5266(03)00085-2

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Gupta, Rupasinghe, Callahan, Natera, Smith, Hill, Roessner and Boughton. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Terpene Synthases as Metabolic Gatekeepers in the Evolution of Plant Terpenoid Chemical Diversity

Prema S. Karunanithi and Philipp Zerbe*

Department of Plant Biology, University of California Davis, Davis, CA, United States

OPEN ACCESS

Edited by:

Kazuki Saito,
RIKEN Center for Sustainable
Resource Science (CSRS), Japan

Reviewed by:

Dae-Kyun Ro,
University of Calgary, Canada
Joseph Chappell,
University of Kentucky,
United States

*Correspondence:

Philipp Zerbe
pzerbe@ucdavis.edu

Specialty section:

This article was submitted to Plant
Metabolism and Chemodiversity,
a section of the journal
Frontiers in Plant Science

Received: 14 June 2019

Accepted: 26 August 2019

Published: 01 October 2019

Citation:

Karunanithi PS and Zerbe P (2019)
Terpene Synthases as Metabolic
Gatekeepers in the Evolution of Plant
Terpenoid Chemical Diversity.
Front. Plant Sci. 10:1166.
doi: 10.3389/fpls.2019.01166

Terpenoids comprise tens of thousands of small molecule natural products that are widely distributed across all domains of life. Plants produce by far the largest array of terpenoids with various roles in development and chemical ecology. Driven by selective pressure to adapt to their specific ecological niche, individual species form only a fraction of the myriad plant terpenoids, typically representing unique metabolite blends. Terpene synthase (TPS) enzymes are the gatekeepers in generating terpenoid diversity by catalyzing complex carbocation-driven cyclization, rearrangement, and elimination reactions that enable the transformation of a few acyclic prenyl diphosphate substrates into a vast chemical library of hydrocarbon and, for a few enzymes, oxygenated terpene scaffolds. The seven currently defined clades (a-h) forming the plant TPS family evolved from ancestral triterpene synthase- and prenyl transferase-type enzymes through repeated events of gene duplication and subsequent loss, gain, or fusion of protein domains and further functional diversification. Lineage-specific expansion of these TPS clades led to variable family sizes that may range from a single TPS gene to families of more than 100 members that may further function as part of modular metabolic networks to maximize the number of possible products. Accompanying gene family expansion, the TPS family shows a profound functional plasticity, where minor active site alterations can dramatically impact product outcome, thus enabling the emergence of new functions with minimal investment in evolving new enzymes. This article reviews current knowledge on the functional diversity and molecular evolution of the plant TPS family that underlies the chemical diversity of bioactive terpenoids across the plant kingdom.

Keywords: terpenoids, terpene synthases, plant specialized metabolism, plant chemical diversity, terpenoid biosynthesis, natural products

INTRODUCTION

Among the wealth of small molecule natural products, terpenoids (also referred to as isoprenoids) form an especially diversified and evolutionary ancient superfamily, which likely emerged alongside the formation of primitive membranes at the very origins of cellular life (Ourisson and Nakatani, 1994). Ubiquitous presence of terpenoids in membranes supports this hypothesis and suggests that ancient archaeobacterial diphitynylglycerol ether membrane components, polyprenols, and derived steranes and sterols represent early terpenoid predecessors (Ourisson and Nakatani, 1994; Rohmer and Bissleret, 1994; Van De Vossenberget al., 1998; Matsumi et al., 2011). From this origin, the staggering diversity of the terpenome has arisen, comprising more than 80,000

compounds (Christianson, 2018) that are widespread across living organisms, including archaea (Matsumi et al., 2011), bacteria (Yamada et al., 2012), fungi (Schmidt-Dannert, 2015), social amoeba (Chen et al., 2016; Chen et al., 2019), marine organisms (Gross and König, 2006), insects (Beran et al., 2016; Lancaster et al., 2018), and plants (Gershenzon and Dudareva, 2007; Tholl, 2015). Plants are the champions of producing different terpenoid structures (Tholl, 2015). This includes a few isoprenoid derivatives with essential roles in plant growth and development such as gibberellins, brassinosteroids, carotenoids, and chlorophylls (Pallardy, 2008; Tripathy and Pattanayak, 2012). Conversely, the vast majority of plant terpenoids represent specialized metabolites that are dedicated to mediating interorganismal interactions or environmental defense and adaptation (Gershenzon and Dudareva, 2007; Tholl, 2015). For example, many terpenoids exhibit potent toxicity and serve as core components of chemical defenses against herbivores, insect pests, and microbial pathogens (Keeling and Bohlmann, 2006a; Vaughan et al., 2013; Schmelz et al., 2014). In addition, functions in allelopathic interactions and roles in abiotic stress responses have been reported (Lopez et al., 2008; Kato-Noguchi and Peters, 2013; Vaughan et al., 2015). Terpenoid bioactivities in cooperative interactions are equally diverse, including various volatile terpenoids essential for attracting pollinators and seed dispersers, as well as in mediating plant–plant and plant–microbe interactions that impact plant fitness (Dudareva and Pichersky, 2000; Pichersky and Gershenzon, 2002; Heil and Ton, 2008; Agrawal and Heil, 2012). Driven by selective pressures to adapt to the biotic and abiotic environments of the ecological niche occupied by individual plant species, specialized terpenoid metabolism has undergone an expansive evolutionary divergence, resulting in often lineage-specific pathways and products (Chen et al., 2011; Tholl, 2015; Zerbe and Bohlmann, 2015). Biosynthesis and accumulation of these compounds also are typically restricted to only a subset of organs, tissues, or developmental stages and may be tightly regulated by internal or external stimuli, granting plants the ability to fine-tune the deployment of terpenoids for mediating dynamic interactions with the environment (Keeling and Bohlmann, 2006a; Tholl, 2006; Schmelz et al., 2014). Owing to their diverse bioactivities, terpenoid-forming plants and their products have a long history of exploitation for human benefit. Historically, large-scale extraction of terpenoid resins from coniferous trees has been a resource for producing “turpentine”—giving the metabolite class its name—and continue to be of economic relevance for the manufacture of biopolymers and inks (Bohlmann and Keeling, 2008). Other uses of plant terpenoids span various industrial sectors, including flavors and fragrances (Lange et al., 2011; Schalk et al., 2012; Philippe et al., 2014; Celedon and Bohlmann, 2016), pharmaceuticals and cosmetics (Paddon et al., 2013; Pateraki et al., 2017; Booth and Bohlmann, 2019; Zager et al., 2019), biofuels (Peralta-Yahya et al., 2012; D’espau et al., 2015), and natural rubber (Oh et al., 2000; Cornish and Xie, 2012; Qu et al., 2015).

The biological and economic relevance of terpenoids has fostered long-standing efforts in understanding the metabolic enzymes that generate terpenoid chemical diversity. Following

common metabolic patterns of scaffold-forming and tailoring reactions in specialized metabolism (Anarat-Cappillino and Sattely, 2014), terpenoid biosynthesis proceeds through conversion of central 5-carbon isoprenoid precursors into a range of core scaffolds that are then functionally elaborated to generate the diversity of terpenoid bioactivities (Davis and Croteau, 2000; Chen et al., 2011; Hamberger and Bak, 2013; Zerbe and Bohlmann, 2015) (**Figure 1**). Functionally distinct enzyme families of terpene synthase (TPS) and cytochrome P450 monooxygenase (P450) enzymes are the major drivers of scaffold formation and functional modifications, respectively (Peters, 2010; Chen et al., 2011; Nelson and Werck-Reichhart, 2011; Zerbe and Bohlmann, 2015; Banerjee and Hamberger, 2018; Bathe and Tissier, 2019) (**Figure 1**). In particular, TPSs serve as the gatekeepers of species-specific terpenoid pathways, catalyzing stereo-specific carbocation cascades that transform a handful of common prenyl diphosphate substrates into the core scaffolds of numerous structurally distinct terpenoid groups. Recent years have witnessed groundbreaking advances in genomics and biochemical tools that have enabled the discovery of TPS and P450 enzymes at an unprecedented scale and can be combined with versatile metabolic engineering approaches toward producing a broader range of terpenoid bioproducts (Kearling, 2012; Kitaoka et al., 2015; Mafu and Zerbe, 2018). Building on comprehensive reviews on terpenoid biological function, regulation, and biochemistry (Dudareva and Pichersky, 2000; Tholl, 2006; Gershenzon and Dudareva, 2007; Hamberger and Bak, 2013; Lange and Turner, 2013; Schmelz et al., 2014; Tholl, 2015), this review focuses on recent advances in the knowledge of terpenoid biosynthesis and the evolutionary divergence of the TPS family.

METABOLIC ORIGIN OF TERPENOID PRECURSORS

Biosynthesis of C₅ Isoprenoid Building Blocks

The metabolic origin of all terpenoids centers around the assembly of multiples of the common C₅ isoprenoid precursor isopentenyl diphosphate (IPP) and its double-bond isomer dimethylallyl diphosphate (DMAPP) (Lange et al., 2000; Christianson, 2008). Unlike most microbial organisms, plants utilize two distinct pathways for producing these building blocks: the acetyl-CoA-derived cytosolic mevalonate (MVA) pathway and the pyruvate-derived plastidial 2-C-methyl-D-erythritol-4-phosphate (MEP) pathway (McGarvey and Croteau, 1995; Hemmerlin et al., 2012) (**Figure 1**). Presence of the MVA pathway in archaea (albeit with some species being devoid of some pathway enzymes) and phylogenetic relatedness of MVA pathway genes across multiple taxa provide evidence that the MVA pathway represents the ancestral isoprenoid-metabolic route that was present in the last common ancestor and has been vertically transmitted to the descendants (Lombard and Moreira, 2011; Matsumi et al., 2011). By contrast, the plastidial MEP pathway was likely acquired through horizontal gene transfer from different bacterial progenitors such as cyanobacteria and various proteobacteria

(Lange et al., 2000). The metabolic expense of retaining two IPP/DMAPP-metabolic pathways in plants holds apparent advantages by enabling broader ability to evolve specialized terpenoid pathways and better control of compartment-specific isoprenoid pools toward MEP-derived mono- and di-terpenoids, carotenoids, plastoquinones and chlorophyll in plastids and MVA-derived sesquiterpenoids, sterols, brassinosteroids, and triterpenoids. This physical separation of downstream pathways has been supported—for example, by genome-wide co-expression studies in *Arabidopsis* that showed minimal interaction between MVA and MEP genes (Wille et al., 2004; Vranova et al., 2013; Rodriguez-Concepcion and Boronat, 2015). In addition, presently known pathway connections appear to be largely negative in nature, where transcriptional activation of MEP genes correlate with the repression of MVA genes and *vice versa* (Ghassemian et al., 2006; Rodriguez-Concepcion and Boronat, 2015). On the other hand, metabolic compensation, for example, of cytosolic sterol biosynthesis through the MEP pathway has been described (Hemmerlin et al., 2003; Laule et al., 2003). Indeed, cross-talk between both pathways *via* exchange of IPP, DMAPP, and C₁₀₋₁₅ prenyl diphosphate intermediates has been demonstrated in several species (Hemmerlin et al., 2003; Laule et al., 2003; Opitz et al., 2014; Mendoza-Poudereux et al., 2015), indicating that the metabolic fate of MEP- and MVA-derived IPP and DMAPP is not as clear cut. For example, isotope-labeling studies demonstrated incorporation of MEP-derived IPP/DMAPP into both mono- and sesqui-terpenoids in snapdragon (*Antirrhinum majus*) and carrot (*Daucus carota*) (Dudareva et al., 2005; Hampel et al., 2005). Similar work in cotton (*Gossypium hirsutum*) showed contribution of the MVA pathway to C₁₀–C₄₀ terpenoid biosynthesis (Opitz et al., 2014).

How the cross-membrane exchange of MEP and MVA intermediates is coordinated also requires further investigation. Transport of IPP and GPP across the plastidial membrane has been observed in isolated plastids (Soler et al., 1993; Bick and Lange, 2003; Flugge and Gao, 2005), but transporters or alternate transfer mechanisms are thus far unknown (Pick and Weber, 2014). Recent studies further illustrated that terpenoid biosynthesis *via* the MVA and MEP pathways is not solely routed through IPP and DMAPP but can involve a pool of the respective isopentenyl and dimethylallyl monophosphates, IP, and DMAP (Henry et al., 2015; Henry et al., 2018). The IP and IPP pools are controlled by two enzymes families, IP kinases and Nudix hydrolases, that catalyze the phosphorylation and hydrolysis of IP and IPP, respectively (Henry et al., 2015; Henry et al., 2018) (**Figure 1**). IP kinases were first discovered in archaea and Chloroflexi as an alternate pathway for isoprenoid biosynthesis (Dellas et al., 2013). More recently, IP kinase homologs were shown to be widely distributed in plant genomes, where they occur alongside the complete set of MVA and MEP pathway genes (Vannice et al., 2014). In *Arabidopsis*, IP kinase was shown to localize at the cytosol and regulate the formation of both MVA- and MEP-derived terpenoids as based on reverse genetic studies (Henry et al., 2015). *AtIPK* knockout in *Arabidopsis* using T-DNA insertion lines caused a significant decrease in the levels of sterols (37–50%) and sesquiterpenes produced (25–31%). Conversely, overexpression of *AtIPK* in transgenic *Nicotiana*

tabacum led to a 3-fold and 2-fold increase in sesquiterpenes and monoterpenes, respectively. Further efforts to understand the formation of IP/DMAP in plants identified a role of Nudix hydrolases, a superfamily of two-domain hydrolases/peptidases broadly found in bacteria, animals, and plants (Bessman et al., 1996; Ogawa et al., 2005; Magnard et al., 2015; Henry et al., 2018). *In vitro* and genetic studies of the two cytosolic Nudix hydrolases in the *Arabidopsis* genome, *AtNudx1* and *AtNudx3*, demonstrated their efficiency in dephosphorylating IPP and DMAPP (Henry et al., 2018). *Arabidopsis* T-DNA insertion gene knock-down and knock-out lines of *AtNudx1* and *AtNudx3* resulted in increased production of sesquiterpenes (28–60%), monoterpenes (148–503%), and sterols (~50%) whereas overexpression of these enzymes in *N. tabacum* resulted in decreased production of monoterpenes (~50%) and sesquiterpenes (57–88%). Although understanding the broader relevance of IP kinase and Nudix hydrolase genes in plant terpenoid metabolism requires further studies, these collective findings highlight the potential of these pathway reactions to possibly function as additional regulatory mechanisms for balancing the IP/DMAP and IPP/DMAPP pools in the biosynthesis of terpenoids and other isoprenoids (Henry et al., 2015; Henry et al., 2018). Given the dramatic impact of modulating IP kinase and Nudix hydrolases gene expression on pathway productivity, combined tailoring of these pathway nodes holds promise for advanced terpenoid pathway engineering.

Biosynthesis of Prenyl Diphosphate Precursors

Downstream of the IPP and DMAPP biosynthesis, prenyl transferases (PTs) catalyze the sequential condensation of isoprenoid units *via* ionization of the allylic diphosphate ester and subsequent rearrangement of the resulting carbocation to generate prenyl diphosphate metabolites of distinct chain length that serve as universal terpenoid precursors (Nagel et al., 2019) (**Figure 1**). Head-to-tail condensation (C₄–C₁ alkylation) reactions lead to C₁₀ (geranyl diphosphate, GPP), C₁₅ (farnesyl diphosphate, FPP), and C₂₀ (geranylgeranyl diphosphate, GGPP) intermediates as precursors in mono-, sesqui-, and di-terpenoid metabolisms, respectively (**Figure 2A**). Notably, dimerization has been shown to be a major factor impacting PT activity and product specificity. For example, GPP synthases from *Mentha piperita*, *A. majus*, and *Clarkia breweri* require formation of a heterodimer of a small and a large subunits for their enzyme function (Burke et al., 1999; Tholl et al., 2004). Interaction of GPP small subunits with GGPP synthases from *Abies grandis* and *Taxus canadensis* were further shown to modify GGPP synthase product specificity in favor of forming shorter C₁₀ chains (Burke and Croteau, 2002). Similarly, interaction of a *cis*-PT with an unusual *cis*-PT-like scaffolding enzyme was shown to be a key function in rubber biosynthesis in lettuce (*Lactuca sativa*) (Qu et al., 2015). As alternative routes to the common head-to-tail condensation reactions, C₃₀ and C₄₀ prenyl diphosphates are formed *via* head-to-head condensation of FPP or GGPP through the activity of squalene synthases or phytoene synthases en route to triterpenoids and carotenoids, respectively (Christianson, 2008). Similarly, catalysis of non-head-to-tail or irregular C_{1'}–2-3 isoprenoid condensation can occur as exemplified by a PT

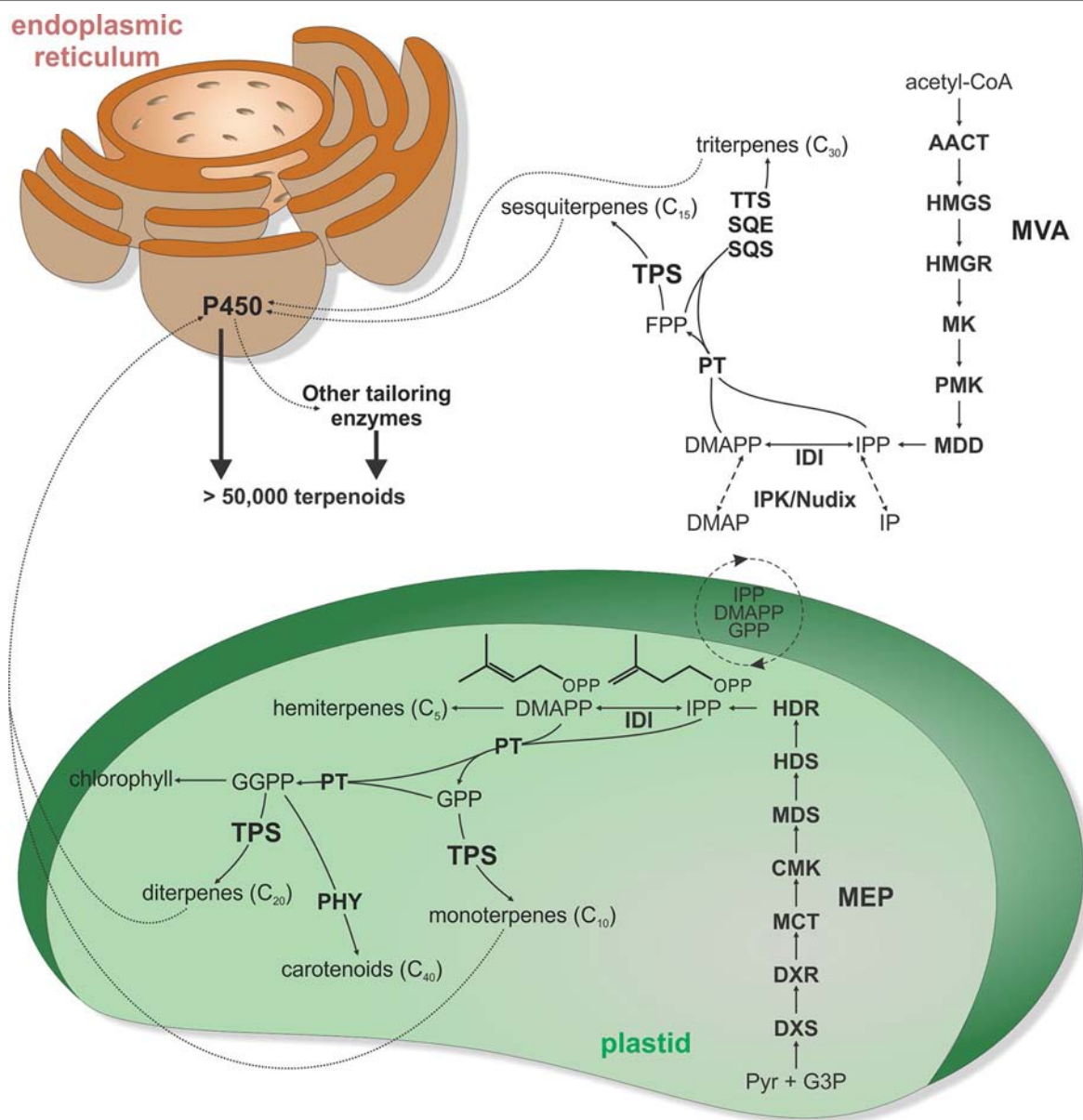


FIGURE 1 | Schematic overview of major terpenoid biosynthetic pathways. All terpenoids are derived from two isomeric 5-carbon precursors, isopentenyl diphosphate (IPP), and dimethylallyl diphosphate (DMAPP). In turn, IPP and DMAPP are formed via two pathways, the cytosolic mevalonate (MVA) pathway originating from acetyl-CoA and the pyruvate and glyceraldehyde-3-phosphate (G3P)-derived 2-C-methyl-D-erythritol-4-phosphate (MEP) pathway located in the plastids. However, active transfers of IPP, DMAPP, GPP, and FPP across the plastidial membrane enable some degree of pathway cross-talk. In addition, interconversion of IPP and DMAPP with their respective monophosphate forms IP and DMAP by IP kinase (IPK) and Nudix hydrolase enzymes can impact pathway flux in terpenoid metabolism. Except for isoprene and hemiterpene (C₅) biosynthesis, condensation of IPP and DMAPP units generates prenyl diphosphate intermediates of different chain length. Condensation of IPP and DMAPP yields geranyl diphosphate (GPP) as the precursor to monoterpenoids (C₁₀), fusing GPP with an additional IPP affords the sesquiterpenoid (C₁₅) precursor farnesyl diphosphate (FPP), and fusing FPP with IPP generates geranylgeranyl diphosphate (GGPP) en route to diterpenoids (C₂₀). Furthermore, condensation of two FPP or two GGPP molecules forms the central substrates of triterpenoid (C₃₀) and carotenoids (C₄₀), respectively. Terpene synthases (TPS) are key gatekeepers in the biosynthesis of C₁₀–C₂₀ terpenoids, catalyzing the committed scaffold-forming conversion of the respective prenyl diphosphate substrates into a range of hydrocarbon or oxygenated structures. These TPS products can then undergo various oxygenations through the activity of cytochrome P450 monooxygenases (P450), followed by further possible functional decorations, ultimately giving rise to more than 80,000 distinct natural products. AACT, acetoacetyl-CoA thiolase; CMK, 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase; DXR, 1-deoxy-D-xylulose 5-phosphate reductase; DXS, 1-deoxy-D-xylulose 5-phosphate synthase; HDR, (E)-4-hydroxy-3-methyl-but-2-enyl diphosphate reductase; HDS, (E)-4-hydroxy-3-methyl-but-2-enyl diphosphate synthase; HMGR, 3-hydroxy-3-methylglutaryl-CoA reductase; HMGS, 3-hydroxy-3-methylglutaryl-CoA synthase; IDI, isopentenyl diphosphate isomerase; MCT, MEP cytidyltransferase; MDD, mevalonate-5-diphosphate decarboxylase; MDS, 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; MK, mevalonate kinase; P450, cytochrome P450-dependent monooxygenase; PHY, phytoene synthase; PMK, phosphomevalonate kinase; PT, prenyl transferase; SCS, squalene synthase; SQE, squalene epoxidase; TPS, terpene synthase; TTS, triterpene synthase.

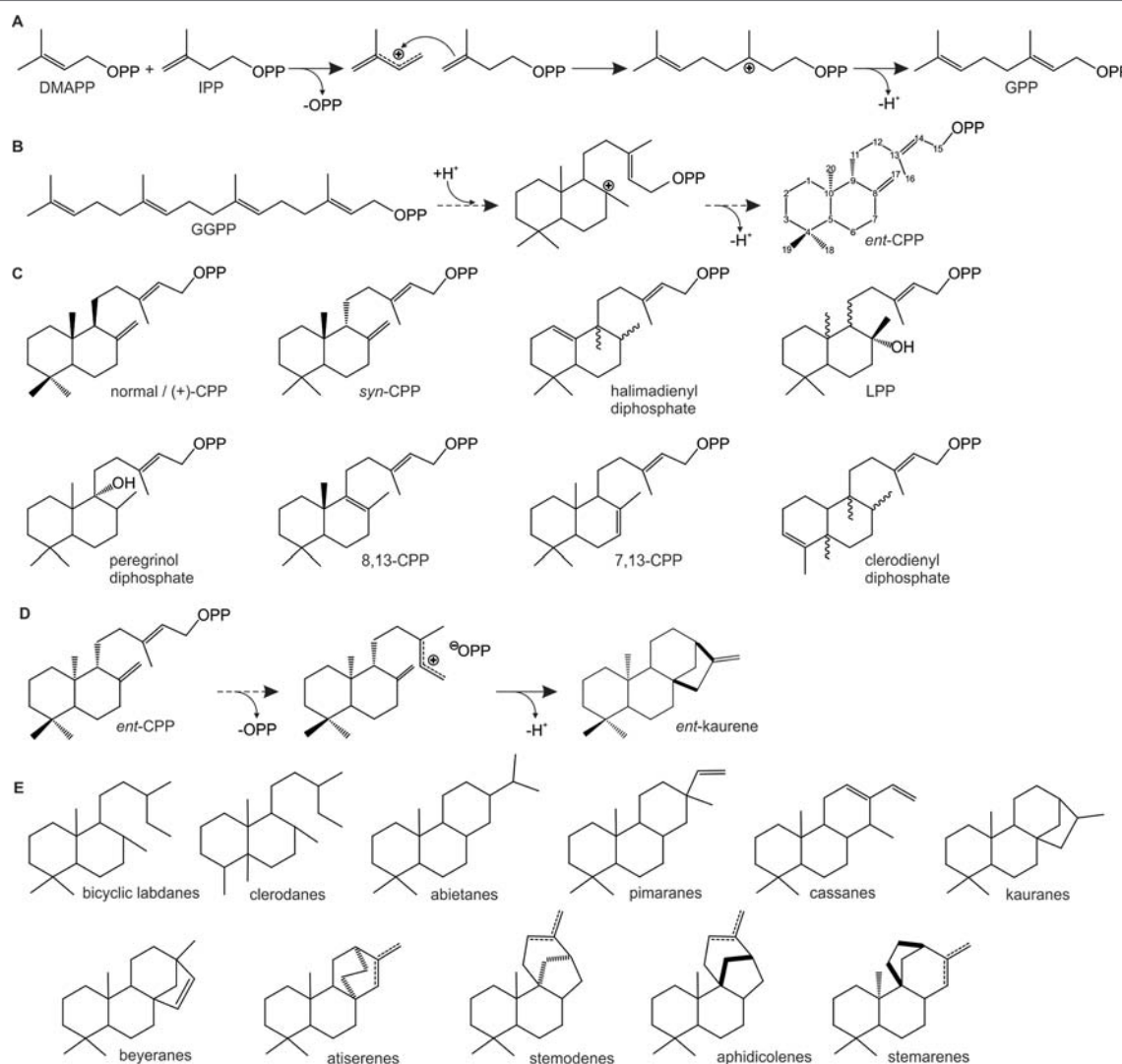


FIGURE 2 | Schematic overview of representative carbocation-driven reactions catalyzed by prenyl transferases (PTs) and diterpene synthases (TPSs). **(A)** PT catalyzed head-to-tail condensation of isopentenyl diphosphate (IPP) with its positional isomer dimethylallyl diphosphate (DMAPP) *via* ionization of the allylic diphosphate ester bond (OPP) and subsequent coupling between the resulting carbocation and the C3–C4 double bond of IPP. Deprotonation of the carbocation intermediate yields geranyl diphosphate (GPP). **(B–C)** Conversion of geranylgeranyl GGPP by class II diTPSs using protonation-initiated cyclization of GGPP to facilitate scaffold rearrangement into bicyclic prenyl diphosphates of *ent*-copalyl diphosphate (*ent*-CPP) **(B)** and related scaffolds of distinct stereochemistry and hydroxylation **(C)**. **(D–E)** Class I diTPS-catalyzed conversion of bicyclic prenyl diphosphate intermediates *via* ionization of the diphosphate moiety and subsequent cyclization and rearrangement through various 1,2-hydride and methyl migrations to form, for example, *ent*-kaurene **(D)** and a range of other labdane diterpene scaffolds **(E)**.

from *Chrysanthemum cinerariaefolium* that forms the irregular monoterpene chrysanthemyl diphosphate (Rivera et al., 2001). Recent years further revealed the biosynthetic enzymes forming plant prenyl diphosphate products of other chain length, including C_{25} intermediates en route to the rare group of defensive sesterterpenoids in *Brassicaceae* and a few other species (Luo et al., 2010; Nagel et al., 2015; Huang et al., 2017; Chen et al., 2019). *Arabidopsis* studies further illustrated a *trans*-type polyprenyl diphosphate synthase with the capacity to produce variable chain length (C_{25} – C_{45}) products (Hsieh et al., 2011).

In addition to product chain length, PT enzymes can be distinguished based on the *cis*- or *trans*-C–C double-bond configuration of their product (Liang et al., 2002). Although PTs

utilize the same isoprenoid substrates, *cis*- and *trans*-type PTs differ in their protein structure and signature motifs that control catalytic specificity (Liang et al., 2002). Similar to class I TPS enzymes (described below), *trans*-PTs feature two aspartate-rich motifs, FARM (DDX_{2/4}D) and SARM (DDX₂D), which are critical for substrate binding, whereas *cis*-PTs lack these motifs, and substrate binding is controlled by Asp and Glu residues broadly distributed within the active site (Liang et al., 2002). Medium- and long-chain ($\geq C_{30}$) PTs more commonly produce *cis*-prenyl diphosphate compounds, most prominently represented by *cis*-PTs of rubber biosynthesis (Asawatreratanakul et al., 2003; Cornish and Xie, 2012; Chakrabarty et al., 2015; Qu et al., 2015). In contrast, the majority of short-chain (C_{10-25})

prenyl diphosphates occur in the *trans* configuration. However, a number of *cis*-PTs have been identified in certain species that produce intermediates featuring *cis*- and *trans*-configured double bonds. For example, short-chain *cis*-PTs were identified in tomato (*Solanum* sp.) (Sallaud et al., 2009; Schilmiller et al., 2009; Akhtar et al., 2013; Matsuba et al., 2015). In the wild tomato variety *Solanum habrochaites*, an FPP synthase, zFPS, was demonstrated to form the *cisoid* FPP form *Z,Z*-FPP, which is further converted by a *Z,Z*-FPP-specific TPS (SBS) to form (+)- α -santalene, (+)-endo- β -bergamotene, and (–)-endo- α -bergamotene (Sallaud et al., 2009). Notably, zFPS localizes to the chloroplast, contrasting the commonly cytosolic localization of *trans*-FPP synthases (Sallaud et al., 2009; Schilmiller et al., 2009; Akhtar et al., 2013). Further gene discovery studies in cultivated tomato (*Solanum lycopersicum*) identified neryl-diphosphate synthase 1 (NDPS1) catalyzing the formation of a *cis*-neryl diphosphate (NPP) as a precursor for a range of monoterpenoids in addition to the canonical *trans*-substrate GPP (Schilmiller et al., 2009; Gutensohn et al., 2013). More recently, a diterpenoid-metabolic cluster was reported in cultivated tomato that included an unusual *cis*-PT (CPT2) that produces the *cisoid* C₂₀ GGPP variant *Z,Z,Z*-nerylneryl diphosphate (NNPP) (Matsuba et al., 2015). NNPP conversion by a TPS (TPS21) and a P450 (CYP71BN1) located within this cluster yielded the unusual diterpenoid lycosantalol (Matsuba et al., 2015). Combinatorial functional analysis of class II TPSs known to convert *transoid* *E,E,E*-GGPP showed a broad capacity to also convert NNPP (Jia and Peters, 2017; Pelot et al., 2019). Together, the identification of *cis*-PT enzymes in an increasing number of species and the capacity of several TPSs to convert both *transoid* and *cisoid* prenyl diphosphate substrates may suggest that *cis*-prenyl diphosphate-derived terpenoids are more widely distributed than previously assumed.

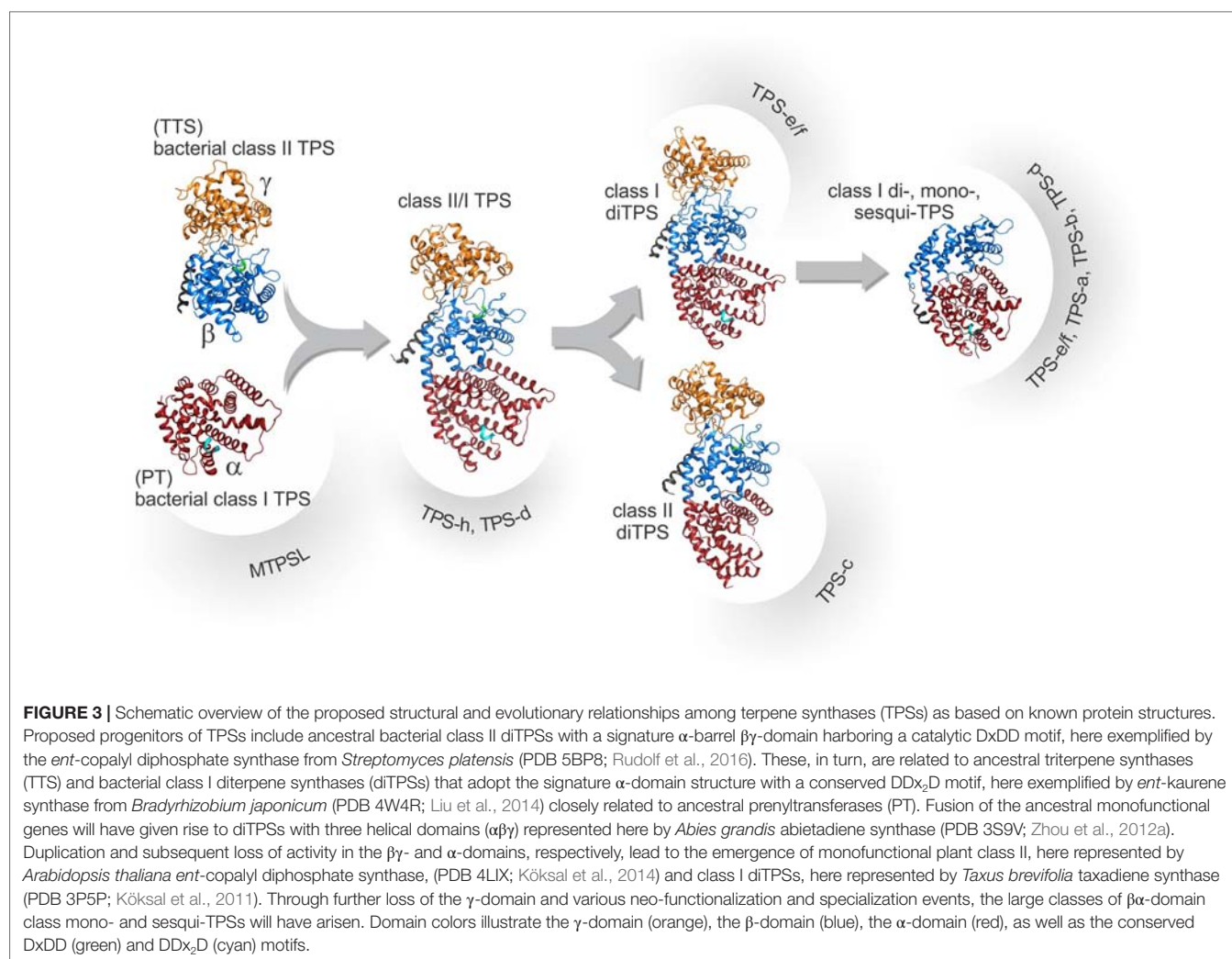
EVOLUTION OF TERPENE SYNTHASES DRIVES TERPENOID CHEMICAL DIVERSITY

The family of TPS enzymes governs the committed scaffold-forming C–C bonding and hybridization reactions in the biosynthesis of terpenoid chemical diversity from a handful of acyclic and achiral C_{5n} prenyl diphosphate substrates. At the core of TPS product specificity is the intramolecular rearrangement of highly reactive carbocation intermediates. Although with far greater variation, these reactions are mechanistically analogous to those observed in PT enzymes (Figure 2), which suggested an evolutionary relationship between both enzyme families (Gao et al., 2012). Structural studies further strengthened this hypothesis by illustrating that these enzymes share a common TPS fold comprised of variations of three conserved helical domains, α , $\beta\alpha$, or $\gamma\beta\alpha$ (Figure 3). Differences in the functionality of these domains distinguish two major classes of TPSs: class II TPSs generate the initial carbocation intermediate *via* substrate protonation and catalyze scaffold rearrangements without cleavage of the diphosphate ester bond, whereas class I TPSs utilize ionization of the diphosphate moiety to form the

intermediary carbocation (Davis and Croteau, 2000; Peters, 2010) (Figures 2B, C). The class II ($\beta\gamma$) domain adopts a characteristic double α -barrel structure that likely evolved from bacterial class II TPSs, which in turn are related to ancestral class II triterpene synthases such as squalene-hopene cyclase (Cao et al., 2010; Gao et al., 2012; Christianson, 2018) (Figure 3). Crystallization of the class II *ent*-copalyl diphosphate (CPP) synthase from *Streptomyces platensis* empirically demonstrated this typical α -barrel $\beta\gamma$ -domain structure featuring a Dx₂E motif closely related to the DxDD signature motif critical for the activity of plant class II diTPSs (Rudolf et al., 2016). Class I activity occurs in the α -helical α -domain (Figure 3), predecessors of which will have been ancestral bacterial class I PT and TPS enzymes, as exemplified by the crystal structure of the class I diTPS *ent*-kaurene synthase from *Bradyrhizobium japonicum* that illustrates the presence of the characteristic α -domain fold along with the signature catalytic DDxxD of class I TPSs (Liu et al., 2014). Such consecutively acting *ent*-CPP and *ent*-kaurene synthases are indeed broadly distributed in plant-associated bacteria, including symbiotic rhizobia such as α -, β -, and γ -proteobacteria (*Rhizobiales*) and some phytopathogens such as species of *Xanthomonas* and *Erwinia* (Morrone et al., 2009; Nagel and Peters, 2017; Nagel et al., 2018). Ancestral *ent*-CPP and *ent*-kaurene synthases are core enzymes in the formation of bioactive gibberellin (GA) phytohormones, and all so far characterized bacterial *ent*-CPP and *ent*-kaurene synthases function as part of GA-biosynthetic operons, albeit with some end product variation ranging from GA precursors to bioactive GA₄ (Nagel and Peters, 2017). Based on the wide distribution of GA-biosynthetic gene clusters in bacteria, it has been suggested that plants acquired the ability to form GAs through ancient events of horizontal gene transfer with soil bacteria (Cao et al., 2010; Smanski et al., 2012), thus providing a selective advantage for phytohormone biosynthesis to control growth and development, as well as a genetic reservoir for the evolution of specialized diterpene metabolites as discussed below.

Distribution of Plant Microbial-Like Terpene Synthases (MTPSLs)

Recent studies revealed that some plant species retained a previously unknown class of microbial-like TPSs, termed microbial TPS-like (MTPSL). A family of MTPSL was first discovered by Li et al. in the *Selaginella moellendorffii* genome, where they co-occur with classical plant TPSs (Li et al., 2012). The 48 MTPSLs identified in *S. moellendorffii* are phylogenetically more closely related to bacterial and fungal TPS-like sequences and differ from classical plant TPSs on the basis of several key features. Firstly, MTPSLs show a distinct gene structure with a higher variability of introns (0–7) as compared to 12–14 introns in classical plant TPSs (Trapp and Croteau, 2001; Li et al., 2012). Secondly, MTPSLs adopt a single domain α -fold closely resembling the structure of microbial α -domain enzymes rather than $\beta\alpha$ -domain plant TPSs (Li et al., 2012). Lastly, alongside the common class I DDx₂D signature motif, MTPSLs contain additional DDx₃D and DDx₃ motifs suggesting a distinct evolutionary origin (Li et al., 2012). Despite their structural distinctiveness, *in vitro* enzyme assays demonstrated that



MTPSLs form common mono- and sesqui-terpene products, including linalool, germacrene D, and nerolidol that naturally occur in *S. moellendorffii* (Li et al., 2012). Following the discovery of MTPSLs in *S. moellendorffii*, members of this TPS class were also identified in the liverwort *Marchantia polymorpha*, the hornwort *Anthoceros punctatus*, the moss *Sphagnum lescurii*, and the monilophyte ferns *Pityrogramma trifoliata* and *Woodsia scopulina*, suggesting a broader distribution across evolutionary older plant lineages (Jia et al., 2016b; Jia et al., 2018b; Xiong et al., 2018) (Figure 4). Functionally active MTPSLs were also identified in the genomes of some red algae (*Rhodophyta*) such as *Laurencia pacifica*, *Porphyridium purpureum*, and *Erythrolobus australicus* (Kersten et al., 2017; Wei et al., 2018). However, expansive genomics studies across several hundred species suggest that MTPSLs are absent in seed plants and green algae (Jia et al., 2016b). Close phylogenetic relationships of MTPSLs with fungal or bacterial TPSs support the evolution of MTPSL genes through multiple events of horizontal gene-transfer events between plants, bacteria, and fungi after the split from the last common ancestor with the green algae lineage (Jia et al., 2016b; Kersten et al., 2017; Jia et al., 2018b; Wei et al., 2018). It can

be speculated that the loss of MTPSLs in seed plant lineages is due to the emergence of sesqui-TPS and mono-TPS functions derived from the ancestral bifunctional diTPSs (Jia et al., 2018b).

Emergence and Diversification of Bifunctional Terpene Synthases

A hallmark event in TPS evolution was the fusion of $\beta\gamma$ - and α -domain enzyme classes that gave rise to bifunctional class II/I diTPSs with a $\beta\gamma\alpha$ -domain architecture (Figure 3), providing an apparent evolutionary advantage of improved metabolite channeling of reactive prenyl diphosphate intermediates. Such bifunctional diTPSs have been identified in fungi, mosses, and gymnosperms (Toyomasu et al., 2000; Keeling and Bohlmann, 2006b; Hayashi et al., 2006; Mafu et al., 2011; Zhou et al., 2012a; Fischer et al., 2015; Kumar et al., 2016) but are absent in angiosperms (Figure 4). Whether such domain fusions occurred in the bacterial donor or after transfer of monofunctional diTPS genes remains to be resolved. Moreover, the relevant horizontal gene transfer events likely included only a subset of genes rather than entire operons. For example, the bryophyte *Physcomitrella patens* contain a single prototypical class II/I

diTPS producing *ent*-kaurene and *ent*-hydroxy-kaurene via *ent*-CPP as an intermediate (Hayashi et al., 2006). However, *P. patens* lacks additional genes required for producing bioactive GAs and instead forms the GA intermediate *ent*-kaurenoic acid that functions as a growth and developmental regulator (Hayashi et al., 2006). Presence of ancestral bifunctional diTPSs involved in the biosynthesis of GA-related compounds in *P. patens* and related land plants (Hayashi et al., 2006), but not algae (Lohr et al., 2012; Jia et al., 2018b; Wei et al., 2018), places the evolutionary origin of plant diterpene metabolism with the emergence of nonseed plant lineages approximately 450 million years ago. Presumably in the same time period, ancestral *ent*-(OH)-kaurene-producing bifunctional diTPSs underwent neo-functionalization toward the biosynthesis of diterpenoids with specialized functions in a number of species. The moss *Hypnum plumaeforme* contains a bifunctional diTPS that forms *syn*-pimara-7,15-diene, a diterpenoid also present in rice as precursor of anti-microbial and allelopathic momilactones (Wilderman et al., 2004; Kato-Noguchi and Peters, 2013; Schmelz et al., 2014; Okada et al., 2016). Similarly, two class II/I diTPSs have been identified in *S. moellendorffii* that produce miltiradiene via the enantiomer of *ent*-CPP (9*R*,10*R*-CPP), namely, normal (9*S*10*S*) CPP or (+)-CPP as an intermediate (*Sm*MDS) and labda-7,13*E*-dien-15-ol (*Sm*CPS/KSL1) via labda-15-en-8-ol diphosphate (LPP) (Mafu et al., 2011; Sugai et al., 2011). Although the physiological relevance of these diterpenoids remains elusive, it is plausible that these compounds or derivatives thereof function in disease and pest defense, considering the bioactivity of closely related metabolites in other plant species (Ma et al., 2012; Helmstädtter, 2013). In contrast, the role of specialized bifunctional class II/I diTPSs of the gymnosperm-specific TPS-d clade is well established, where these enzymes form the abietane- and pimarane-type labdane scaffolds in the biosynthesis of diterpene resin acids (DRAs) that serve as a durable defense against insect pests and associated fungal pathogens (Keeling and Bohlmann, 2006a; Bohlmann, 2011). Bifunctional abietane- and pimarane-type diTPSs were identified in *Ginkgo biloba* (Schepmann et al., 2001), species of fir (*Abies*) (Peters et al., 2000; Zerbe et al., 2012b), spruce (*Picea*) (Martin et al., 2004), and pine (*Pinus*) (Ro and Bohlmann, 2006; Hall et al., 2013). Notably, all so far identified enzymes utilize the enantiomer of *ent*-CPP (9*R*,10*R*-CPP), namely, normal (9*S*10*S*) CPP or (+)-CPP, as an intermediate en route to the their individual diterpene products. In addition, some species feature diTPSs that have undergone additional neo-functionalization. For example, balsam fir (*Abies balsamea*) contains a *cis*-abienol synthase (AbCAS) that catalyzes conversion of GGPP into *cis*-abienol via the C-8 hydroxylated CPP intermediate LPP also observed in *S. moellendorffii* (Mafu et al., 2011; Zerbe et al., 2012b).

The bifunctional diTPSs involved in diterpenoid metabolism of mosses, lycophytes, and gymnosperms are strikingly similar to those identified in fungi, especially in *Ascomycota* and some *Basidiomycota* species where they function in pathogenic or plant growth-promoting pathways (Bomke and Tudzynski, 2009; Quin et al., 2014). A bifunctional diTPS with *ent*-CPP/*ent*-kaurene synthase activity was first identified as a part of a GA-biosynthetic gene cluster in *Gibberella fujikuroi* (genus *Fusarium*), the causal agent of to bakanae disease in rice

(*Oryza sativa*) (Tudzynski and Holter, 1998; Toyomasu et al., 2000; Tudzynski, 2005). Beyond GA-biosynthetic diTPSs, certain fungal species also contain enzymes for producing specialized diterpenoids, as exemplified by the bifunctional aphidicolan-16 β -ol synthase producing a key precursor to aphidicolin toxins in the pathogenic fungus *Phoma betae* (Toyomasu et al., 2004) or the *ent*-CPP/*ent*-kaurene synthase homologs PaDC1/2 involved in the biosynthesis of phyllocladan-triol in *Phomopsis amygdali* (Toyomasu et al., 2008). Recent studies proposed horizontal gene transfer from a plant to an ancestral *Ascomycota* fungus as the primary mechanism underlying the emergence of fungal class II/I diTPSs (Fischer et al., 2015). This hypothesis is supported by several lines of evidence, including the abundance of mutualistic plant-fungal interactions often involving species containing TPS genes such as *Fusarium*, the presence of diTPS in only some *Ascomycota* and *Basidiomycota* species, and the lack of correlation between the phylogenetic relationships of fungal diTPS and the fungal species containing these genes (Fischer et al., 2015). Interestingly, domain fusions between diterpenoid-biosynthetic enzymes are not limited to bifunctional class II/I diTPSs in fungi and other species but also include other chimeric enzymes that, for example, represent fusions of PT and TPS domains (Minami et al., 2018; Mitsuhashi and Abe, 2018), such as the *P. amygdali* Fusicoccadiene synthase that contains an N-terminal class I TPS domain and a C-terminal PT domain and is involved in the biosynthesis of *Fusicoccin* toxins (Toyomasu et al., 2007).

Functional Radiation of Monofunctional Terpene Synthases

The absence of bifunctional class II/I diTPSs in angiosperms (Chen et al., 2011) highlights another milestone in the expansion of the TPS family; the duplication and sub-functionalization of ancestral bifunctional $\gamma\beta\alpha$ -domain diTPSs with one descendent retaining class II (*ent*-CPP synthase) activity in the $\beta\alpha$ -domain and the other copy acting as a monofunctional class I diTPS (*ent*-kaurene synthase) with a functional α -domain (Figures 2B, C and 3). Early examples of such monofunctional enzymes have been described in *S. moellendorffii*, the liverwort *M. polymorpha* (Li et al., 2012; Kumar et al., 2016), and gymnosperms (Keeling et al., 2010) (Figure 4). While ancient vascular plants such as *S. moellendorffii* did not yet use these enzymes for producing bioactive GAs (Aya et al., 2011), monofunctional *ent*-CPP and *ent*-kaurene synthase activities in GA biosynthesis are conserved across vascular plants (Keeling et al., 2010; Chen et al., 2011) (Figures 2B, D). In addition to their critical role in phytohormone metabolism, monofunctional *ent*-CPP, and *ent*-kaurene synthases will have served as a major genetic reservoir for the lineage-specific expansion of functionally diverse class II and class I TPSs across the plant kingdom (Zi et al., 2014) (Figure 4).

Derived from ancestral *ent*-CPP synthases, the TPS-c clade of class II diTPSs has undergone a relatively moderate diversification with known enzymes differing predominantly in their expression patterns and product-specificity toward a range of alternate stereo- and double-bond isomers and hydroxylated intermediates (Peters, 2010; Zerbe and Bohlmann, 2015) (Figures 2B, C). Most prominently, the *ent*-CPP enantiomer (+)-CPP, first identified as an intermediate of gymnosperm class

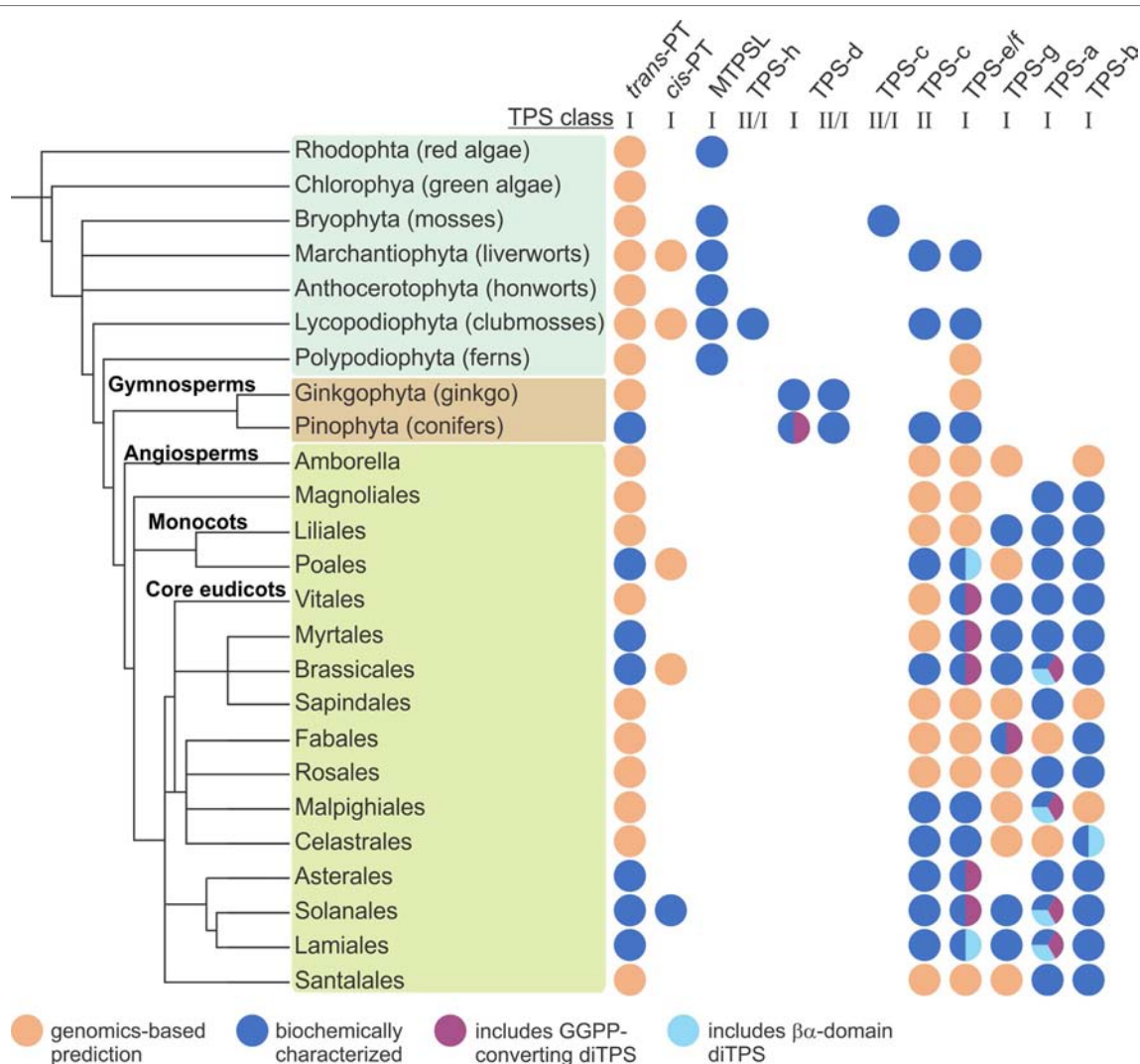


FIGURE 4 | Distribution of terpene-biosynthetic enzyme families across major orders across the plant kingdom. PT, prenyltransferases; MTPSL microbial-type terpene synthases (TPSs); TPS-h, *Selaginella*-specific bifunctional class II/I diterpene synthases (diTPSs); TPS-d, gymnosperm-specific class I mono-, sesqui-, di-TPSs, and bifunctional class II/I diTPSs; TPS-c, monofunctional class II diTPSs; TPS-e/f, monofunctional class I diTPSs; TPS-g, monofunctional class I mono-, sesqui-, and di-TPSs; TPS-a, monofunctional class I sesqui- and di-TPSs; TPS-b, monofunctional class I mono-TPSs.

II/I diTPSs (Peters et al., 2000; Martin et al., 2004), is also a core precursor to many labdane-related specialized diterpenoids in various *Lamiaceae* species (Ma et al., 2012; Brückner et al., 2014; Gao et al., 2014; Zerbe et al., 2014; Božić et al., 2015; Cui et al., 2015; Ignea et al., 2016; Scheler et al., 2016) and some *Poaceae* crops such as maize (*Zea mays*) and wheat (*Triticum aestivum*) (Wu et al., 2012; Mafu et al., 2018). Class II diTPSs forming the alternate stereoisomer *syn*-CPP (9*S*,10*R*-CPP) appear to be of narrower taxonomic distribution with current examples limited to some *Poaceae* grasses (Otomo et al., 2004; Xu et al., 2004; Pelot et al., 2018). While *ent*-CPP, (+)-CPP, and *syn*-CPP are the most commonly occurring labdane diterpene precursors, variable series of 1,2-methyl and/or hydride migrations prior to carbocation neutralization can result in other isomeric structures (Peters, 2010). Examples include clerodienyl diphosphate synthases identified in phylogenetically distant plants such as

the *Lamiaceae* *Salvia divinorum*, the *Poaceae* grass switchgrass (*Panicum virgatum*), and *Celastraceae* *Tripterygium wilfordii* (Pelot et al., 2016; Chen et al., 2017; Hansen et al., 2017a; Pelot et al., 2018); 7,13-CPP synthases described in *S. moellendorffii* and the *Asteraceae* *Grindelia robusta* (Mafu et al., 2011; Zerbe et al., 2015); and most recently, 8,13-CPP synthases in maize and switchgrass (Murphy et al., 2018; Pelot et al., 2018). In addition to variations in the scaffold rearrangement, a few class II diTPSs evolved the ability to terminate the carbocation *via* oxygenation rather than deprotonation, a function already present in the *A. balsamea* *cis*-abienol synthase that forms LPP as an intermediate (Zerbe et al., 2012b). Here, oxygenation commonly occurs at the C-8 position to yield LPP with several such enzymes known (Falara et al., 2010; Caniard et al., 2012; Zerbe et al., 2013; Pelot et al., 2017). But also a C-9-oxygenated product has been observed that is formed likely *via* alternate 1,2-hydride shift

between C-8 and the neighboring methine group of the labda-13-en-8-yl diphosphate carbocation prior to water quenching (Zerbe et al., 2014; Heskes et al., 2018).

While the functional diversification of class II diTPSs has been largely limited to variations in product specificity, the family of class I diTPSs has seen a vast expansion and functional radiation through which the large clades of class I diTPS (gymnosperm TPS-d, angiosperms TPS-e/f), and ultimately, sesqui-TPSs (gymnosperm TPS-d, angiosperm TPS-a) and mono-TPSs (gymnosperm TPS-d, angiosperm TPS-b) have arisen (Chen et al., 2011; Gao et al., 2012) (Figure 3). In angiosperms, the predominant blueprint for this class I TPS expansion will have been the repeated duplication and functionalization of ancestral monofunctional *ent*-kaurene synthases within the TPS-e/f clade (Peters, 2010; Zi et al., 2014; Zerbe and Bohlmann, 2015). Beyond the divergence of enzyme products, many specialized class I diTPSs exhibit broad promiscuity for converting different class II diTPS intermediates into the large class of labdane-related diterpenoids (Figure 2E). This biochemical capacity enables modular pathway networks where functionally distinct enzymes can act in different combinations to generate a wider spectrum of possible products. It then appears that dividing class II and class I activities into two monofunctional enzymes has provided an evolutionary advantage over the improved intermediate channeling in bifunctional diTPSs that emerged through the ancestral domain fusion events. Numerous examples of species-specific modular diterpenoid-metabolic networks have been described, including the biosynthesis of stress defensive diterpenoid networks in several *Poaceae* crops such as wheat, rice, and maize (Xu et al., 2007a; Morrone et al., 2011; Zhou et al., 2012b; Zerbe et al., 2014; Cui et al., 2015; Fu et al., 2016; Mafu et al., 2018), as well as specialized diterpenoid metabolism in species of *Salvia* and other *Lamiaceae* (Brückner et al., 2014; Zerbe et al., 2014; Cui et al., 2015; Heskes et al., 2018). Further studies on how interconnected pathway branches are regulated will be essential to better understand how metabolic flux is coordinated between general and specialized pathways, as well as specialized pathway branches sharing key intermediates. Unlike angiosperms where such modular pathways of pairwise-acting monofunctional class II and class I diTPSs are the major metabolic strategy (Peters, 2010; Morrone et al., 2011; Mafu et al., 2015; Zerbe and Bohlmann, 2015), specialized diterpenoid metabolism in gymnosperms largely relies on ancestral bifunctional enzymes. However, the existence of modular pathways also in conifers was suggested by the discovery of monofunctional class I diTPSs in jack pine (*Pinus banksiana*) and lodgepole pine (*Pinus contorta*) that derived from bifunctional progenitors and are capable of utilizing the (+)-CPP intermediate produced by bifunctional class II/I enzymes to form a set of pimarane-type labdanes (Hall et al., 2013). Likewise, two groups of predictably monofunctional class I diTPSs that likely evolved from both TPS-d and TPS-e/f diTPSs were recently identified in Western red cedar (*Thuja plicata*) (Shalev et al., 2018). Biochemical analysis of these enzymes may shed light on the distribution of modular class II-class I diTPS reactions in gymnosperm specialized diterpenoid metabolism.

This expansive diversification of class I diTPSs resulted in a multitude of enzymes with altered substrate/product specificity

and/or gene expression patterns as a critical contribution to the emergence of species-specific specialized functions (Peters, 2010; Zerbe and Bohlmann, 2015). Mostly, these catalytic alterations resulted in variations of the carbocation rearrangement to yield different pimarane, abietane, clerodane, kaurane, dolabradane, and related labdane scaffolds (Figure 2E). While the majority of these diTPS products are hydrocarbon scaffolds, a few class I diTPSs were discovered in recent years that terminate the intermediary carbocation not by deprotonation but *via* hydroxylation in a manner analogous to class II diTPSs that produce C-8 or C-9 hydroxylated prenyl diphosphate products (Peters, 2010; Zerbe and Bohlmann, 2015). Although examples are presently rare, such enzymes seem to be widely distributed across different plant families, including *P. patens* class I/II 16 α -hydroxy-kaurane synthase (Hayashi et al., 2006), *S. moellendorffii* labda-7,13E-dien-15-ol synthase (Mafu et al., 2011), 13-hydroxy-8(14)-abietene synthases in several gymnosperm species (Keeling et al., 2011), 16 α -hydroxy-*ent*-kaurane synthases from *T. wilfordii* (Hansen et al., 2017a) and *Populus trichocarpa* (Irmisch et al., 2015), nezukol synthase from *Isodon rubescens* (Pelot et al., 2017), and *Salvia sclarea* sclareol synthase (Caniard et al., 2012). The regio-specific hydroxylation reactions catalyzed by these diTPSs suggests that the ability of TPSs to ligate a water molecule or hydroxyl group in the nonpolar active site for coordinated carbocation quenching emerged multiple times independently in terpenoid evolution. Notably, sclareol synthase belongs to a recently discovered group of $\beta\alpha$ -bi-domain class I diTPSs that have undergone loss of the γ -domain (Figure 3). Members of this diTPS group form a separate branch in the TPS-e/f clade and have so far been identified in *Poaceae* and *Lamiaceae* species (Figure 4), where they almost invariably form specialized terpenoid products and exhibit broad substrate promiscuity, in some cases, such as wheat (*T. aestivum*) KSL5 and maize TPS1 spanning mono-, sesqui-, and di-terpenoid products (Hillwig et al., 2011; Caniard et al., 2012; Zerbe et al., 2013; Zerbe et al., 2014; Fu et al., 2016; Jia et al., 2016a; Pelot et al., 2016; Pelot et al., 2017; Pelot et al., 2018). Close phylogenetic relationships to TPS-e/f diTPSs combined with partial *ent*-kaurene synthase activity of a few enzymes suggest that these $\beta\alpha$ -domain diTPSs derived more recently from $\gamma\beta\alpha$ -domain class I enzymes. These $\beta\alpha$ -bi-domain diTPSs resemble the likely progenitors of not only specialized class I diTPSs but the vast classes of modern $\beta\alpha$ -domain mono- and sesqui-TPSs. Here, loss of the γ -domain will have been accompanied by various active site modifications toward converting the shorter C₁₀ and C₁₅ substrates and catalyzing manifold distinct scaffold rearrangements and, in case of the sesqui-TPS family, alteration of enzyme subcellular localization through loss of the N-terminal plastidial transit peptide. Unlike the typically mid-sized diTPS families of 2–30 members, mono- and sesqui-TPS families have undergone a far greater expansion that resulted in diverse families of—for example, 69 TPSs in grape (*Vitis vinifera*) (Martin et al., 2010) and 113 TPS genes in *Eucalyptus* (Külheim et al., 2015), enabling these species to produce an astounding variety of smaller and more volatile terpenoids.

In addition to diTPS converting class II enzyme products, class I diTPSs with the ability to directly convert GGPP as a substrate emerged multiple times during terpenoid evolution

and are abundant in a few plant families (Figure 4). In gymnosperms, known GGPP-converting monofunctional class I diTPSs are currently limited to taxadiene synthases in species of yew (*Taxus* spp.) that produce the taxane scaffold in the biosynthesis of the chemotherapeutic agent Taxol (Williams et al., 2000) and pseudolaratriene synthase from golden larch (*Pseudolarix amabilis*) that forms an unusual 5–7–ring scaffold en route to the bioactive anti-cancer compound pseudolaric acid B (Mafu et al., 2017) (Figure 5). These enzymes adopt the

characteristic $\gamma\beta\alpha$ -domain architecture of gymnosperm diTPSs and phylogenetic analyses clearly indicate that they represent descendants of bifunctional class II/I diTPSs rather than mono- or sesqui-TPSs of the TPS-d clade. In angiosperms, the evolutionary path leading to GGPP-converting diTPSs appears to be more diverse and resulted in enzymes that can produce linear, unusual polycyclic, and macrocyclic scaffolds (Kirby et al., 2010; Vaughan et al., 2013; Falara et al., 2014; King et al., 2014; Luo et al., 2016) (Figure 5).

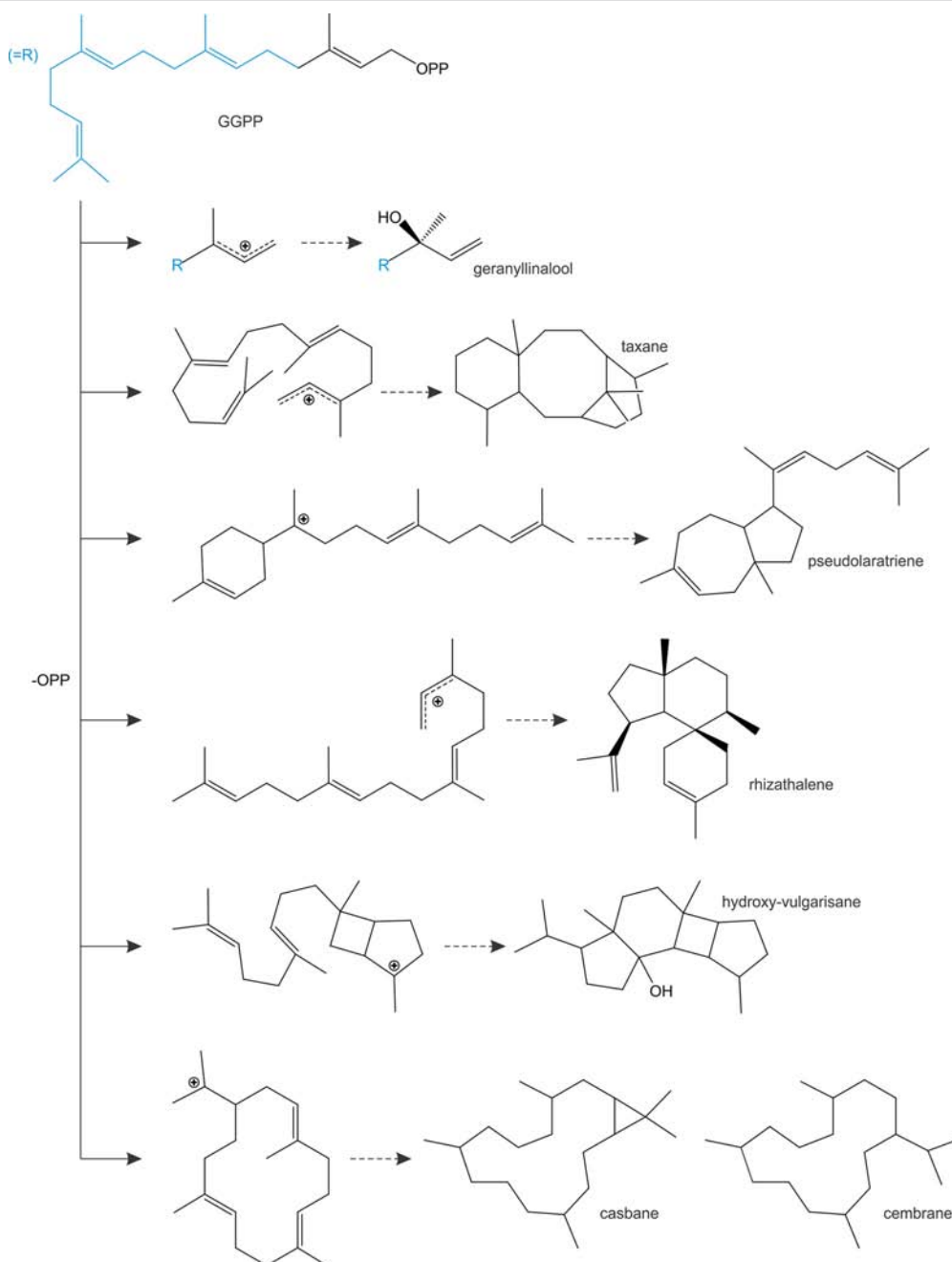


FIGURE 5 | Schematic overview of prominent linear, polycyclic and macrocyclic diterpene scaffolds that are formed through the activity of monofunctional class I diTPSs that are capable of directly converting geranylgeranyl diphosphate (GGPP) as a substrate.

Among these class I diTPSs, geranylinalool synthases (GLSs) are $\gamma\beta\alpha$ -domain diTPSs that uniquely catalyze the ionization of the GGPP diphosphate ester bond without subsequent cyclization to occur (Figure 5), resulting in an acyclic geranylinalool intermediate in the biosynthesis of the homoterpenoid (*E,E*)-4,8,12-trimethyltrideca-1,3,7,11-tetraene (TMTT) with anti-herbivory activity (Falara et al., 2014). Intriguingly, most of the known GLSs form an ancient branch within the TPS-e/f clade of kaurene(-like) synthases, whereas enzymes from *Fabaceae* species fall into the more divergent TPS-g clade mostly comprised of unusual mono- and sesqui-TPSs (Chen et al., 2011; Falara et al., 2014). Although currently known GLSs are restricted to angiosperm species, phylogenetic studies suggest that TPS-e/f GLSs represent an ancient diTPS family that arose from a common ancestor predating the split of the gymnosperm and angiosperm lineages (Falara et al., 2014). Conversely, *Fabaceae*-specific GLSs of the TPS-g family are likely the result of more recent independent evolutionary events. Alike GLSs, class I enzymes converting GGPP into non-labdane poly- or macro-cyclic scaffolds have been identified in a few angiosperm families (Figure 4), including casbene synthases and related macrocyclases in *Euphorbiaceae* (Mau and West, 1994; Kirby et al., 2010; King et al., 2014; Luo et al., 2016), cembratrienol synthases in *Solanaceae* species (Ennajdaoui et al., 2010), *Arabidopsis* rhizathalene synthase (Vaughan et al., 2013), and most recently, 11-hydroxy vulgarisane synthase from the *Lamiaceae* *Prunella vulgaris* (Johnson et al., 2019b) (Figure 5). All known members of this group represent $\beta\alpha$ -bi-domain class I diTPSs and do not belong to the broad TPS-e/f family of class I diTPSs but instead form a distinct branch in the TPS-a clade of sesqui-TPSs (Kirby et al., 2010; King et al., 2014; Johnson et al., 2019b). Two possible evolutionary routes toward these more unusual diTPSs can be envisioned: evolution from TPS-e/f type diTPSs through neo-functionalization and loss of the plastidial signaling peptide or divergence from sesqui-TPS progenitors, which would have involved the re-acquisition of a transit peptide that was previously lost in the evolution of the cytosolic sesqui-TPS family. The latter hypothesis is supported by the close phylogenetic relationship of TPS-a diTPSs and sesqui-TPSs of the TPS-a clade (King et al., 2014; Luo et al., 2016; Johnson et al., 2019b). Such re-evolution events toward diterpenoid-producing TPSs appear not be restricted to the TPS-a clade, since a monofunctional class I diTPS-producing multiradiene was identified in *T. wilfordii* that clusters with the TPS-b clade of mono-TPSs rather than with other multiradiene synthases of the TPS-e/f clade (Hansen et al., 2017a). These collective findings support a highly branched rather than linear evolutionary diversification of TPS functions with many such bifurcations still unknown.

CATALYTIC PLASTICITY OF PLANT TERPENE SYNTHASES

The rapid evolutionary divergence of terpenoid metabolism is aided by the extensive functional plasticity of TPSs. Hence, there has been a long-standing interest in deciphering the mechanisms underlying TPS catalysis and substrate/product specificity. Although TPS structural studies have provided important

insight into TPS catalysis, successful crystallization of plant TPSs currently remains limited to a handful of examples, including *Arabidopsis* *ent*-CPP synthase (class II diTPS) (Köksal et al., 2014), *Taxus brevifolia* taxadiene synthase (class I diTPS) (Köksal et al., 2011), *A. grandis* abietadiene synthase (class II/I diTPS) (Zhou et al., 2012a), and a few mono- and sesqui-TPS enzymes (Starks, 1997; Whittington et al., 2002; Hyatt et al., 2007; Shishova et al., 2008; Gennadios et al., 2009; McAndrew et al., 2011). The conserved nature of the TPS fold further has enabled numerous homology-based structure-function studies that provided a deeper insight into the relative ease of functional change in TPS enzymes, where as little as a single residue mutation can alter the active site contour that largely determines product specificity (for detailed reviews see also Gao et al., 2012; Christianson, 2017).

Early work on the class II/I abietadiene synthase from *A. grandis* demonstrated that class II catalysis uses a general acid-base mechanism to bring about the cyclo-isomerization of GGPP, whereby the acid function is provided by the middle aspartate of the conserved DxDD motif (Peters and Croteau, 2002; Priscic et al., 2007; Peters, 2010). Molecular-level insight into this mechanism was recently gained through solving of the crystal structure of *Arabidopsis* *ent*-CPP synthase at high resolution (Köksal et al., 2014). In this study, Köksal and coworkers elegantly demonstrated that proton transfer with this conserved aspartate is enabled by hydrogen-bonded proton wires that link the active site to the bulk solvent (Köksal et al., 2014). Accompanying mutagenesis studies identified the relevant catalytic base in *Arabidopsis* *ent*-CPP synthase as a water molecule that, in turn, is coordinated by a dyad of two histidine and asparagine residues conserved in *ent*-CPP synthases (Mann et al., 2010; Potter et al., 2014) (Figure 6A). Alanine substitution of these residues resulted in water quenching rather than deprotonation of the labd-enyl carbocation intermediate to yield hydroxylated *ent*-8-hydroxy-CPP products (Potter et al., 2014; Mafu et al., 2015). The widely conserved relevance of this catalytic dyad has been further supported by site-directed mutagenesis of analogous residues in the bifunctional *ent*-CPP/*ent*-kaurene synthase from *P. patens* and the fungus *Fusarium fujikuroi*, which both resulted in the formation of *ent*-LPP as the class II product (Kawaide et al., 2011; Mafu et al., 2015). Likewise, in the class II/I diTPS, abietadiene synthase from *A. grandis* substitution of a tyrosine corresponding to the catalytic histidine and a nearby histidine led to the redirection of class II specificity from (+)-CPP to 8 α -hydroxy-CPP (Criswell et al., 2012). In addition to its critical role in *ent*-CPP synthase activity, residues in the position of this dyad have been shown to impact product specificity in several class II diTPSs. For example, substitution of the *Arabidopsis* *ent*-CPP synthase histidine with phenylalanine or tyrosine rather than Alanine redirected product outcome toward the formation of clerodienyl diphosphate (Potter et al., 2016b) (Figure 6A). Strikingly, reciprocal mutagenesis of the corresponding active site positions in recently discovered clerodienyl diphosphate synthases from *S. divinorum* and *T. wilfordii* blocked the series of migration reactions required for forming the clerodane scaffold and resulted in premature deprotonation to form *ent*-CPP (Pelot et al., 2016; Hansen et al., 2017b). Further supporting this evidence, a recent study by Schulte et al. showed

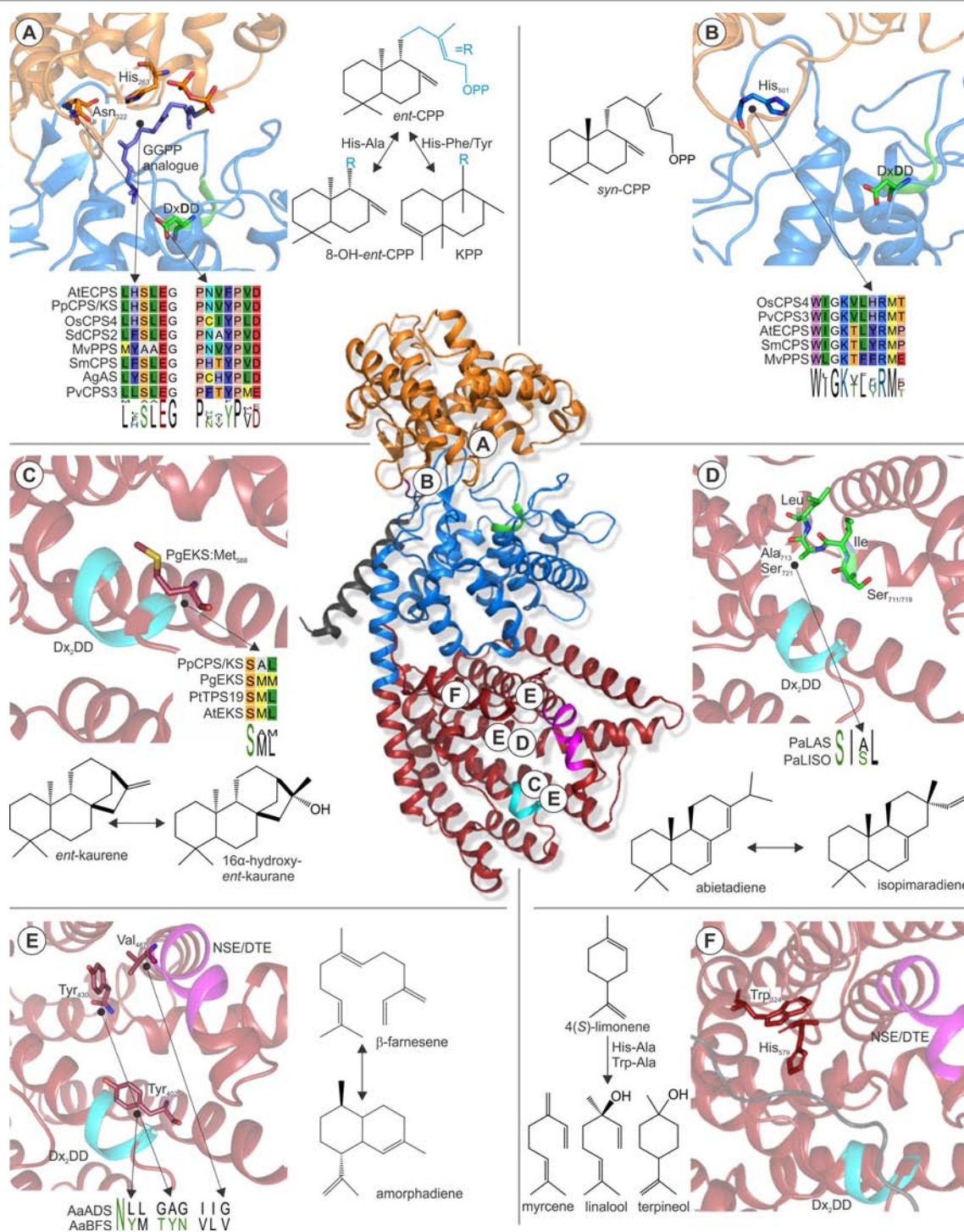


FIGURE 6 | Continued

that the catalytic dyad in a conserved clade of *Lamiaceae* (+)-CPP synthases is represented by a hydrogen-bonded histidine-tyrosine pair (Schulte et al., 2018). Mutagenesis especially of the tyrosine position in *Salvia miltiorrhiza* (+)-CPP synthase as well as other *Lamiaceae* class II diTPSs showed a dramatic impact on product outcome. Likewise, alanine substitution of

a corresponding phenylalanine residue in a 8,13-CPP synthase from switchgrass (PvCPS3) resulted in both positional isomers and hydroxylated forms of the native 8,13-CPP product (Pelot et al., 2018), thus highlighting the relevance of these active site positions in specialized class II diTPSs across diverse plant species.

FIGURE 6 | Examples of active site residues with impact on class II or class I TPS product specificity. **(Center)** Structure of *Abies grandis* abietadiene synthase (PDB 3S9V; Zhou et al., 2012a) that adopts the prototypical α -helical TPS fold with variations in three domains γ (orange), β (blue), and α (red). Relative locations of the highlighted active site residues are indicated **A–F**. **(A)** A widely conserved His-Asn dyad is critical for stereo-specificity of *Arabidopsis ent*-CPP synthase (PDB 4LIX; Köksal et al., 2014) and other *ent*-CPP synthases (Mann et al., 2010; Potter et al., 2014; Mafu et al., 2015). Substitution of this dyad can result in the formation of the alternate clerodienyl diphosphate product (Potter et al., 2014; Pelot et al., 2016). **(B)** His501 in the rice *syn*-CPP synthase OsCPS4 is critical for the stereo-specific formation of *syn*-CPP and is conserved in known *syn*-CPP synthases but not class II diTPS producing alternate CPP stereoisomers (Potter et al., 2016a; Pelot et al., 2018). **(C)** A conserved Met residue in *ent*-kaurene synthases from *Picea glauca* and other species was shown to control *ent*-kaurene formation (Xu et al., 2007b; Zerbe et al., 2012a). **(D)** Mutational studies of corresponding Ser-Ile-Ala-Leu and Ser-Ile-Ser-Leu motifs located at the hinge region of helix G1/2 of *P. abies* levopimaradiene/abietadiene and isopimaradiene synthase showed their critical role in producing abietane or pimarane scaffolds (Keeling et al., 2008). **(E)** Three residues were identified in the active site of *Artemisia annua* β -farnesene synthase, reciprocal exchange of which to corresponding residues in *A. annua* amorphadiene synthase that control activation (Tyr402), reversion (Tyr430), and restoration (Val476) of cyclization capacity (Salmon et al., 2015). **(F)** Two residues, Trp324 and His579, were shown in limonene synthase of *Mentha spicata* to control the reactions cascade toward the natural product 4(S)-limonene (Srividya et al., 2015). The signature catalytic motifs of the class II (DxDD, green) and class I (DDx₂D, cyan; NSE/DTE, magenta) active sites are highlighted. Protein abbreviations: AtECPS, *Arabidopsis thaliana ent*-copalyl diphosphate (CPP) synthase; OsCPS4, *Oryza sativa syn*-CPP synthase; PvCPS3, *Panicum virgatum* 8,13-CPP synthase; SmCPS, *Salvia miltiorrhiza* (+)-CPP synthase; MvCPS1, *Marrubium vulgare* peregriol diphosphate synthase; SdCPS2, *Salvia divinorum* clerodienyl diphosphate (KPP) synthase; PpCPS/KS, *Physcomitrella patens* CPP/*ent*-kaurene synthase; AgAS, *Abies grandis* abietadiene synthase; PgEKS, *Picea glauca ent*-kaurene synthase; PtTPS19, *Populus trichocarpa ent*-kaurene synthase; PaLAS, *Picea abies* levopimaradiene/abietadiene synthase; PaISO, *P. abies* isopimaradiene synthase; AaBFS, *Artemisia annua* β -farnesene synthase; AaADS, *A. annua* amorphadiene synthase.

Another key position contributing to product specificity in class II diTPSs was identified as a histidine (His501) residue in the rice *syn*-CPP synthase OsCPS4, where mutagenesis to aspartate or phenylalanine resulted in additional scaffold rearrangements to form *syn*-halimadienyl diphosphate (Potter et al., 2016a) (**Figure 6B**). A later study on the product specificity of the functionally unique peregriol diphosphate synthase from *Marrubium vulgare* (MvCPS1) showed that substitution of the corresponding phenylalanine residue and a proximal tryptophan in MvCPS1 also redirected product outcome to yield a halimadienyl diphosphate scaffold (Mafu et al., 2016). These structure-function studies in conjunction with the conservation of the relevant histidine residue in known *syn*-CPP synthases, but not functionally distinct class II diTPSs (**Figure 6B**), support the relevance of this position for controlling biosynthesis of the *syn*-CPP stereoisomer (Potter et al., 2016a). Given the relatively smaller functional range of plant class II diTPSs, knowledge of active site determinants controlling product specificity can facilitate sequence-based prediction of class II diTPS functions as additional residues and functionally distinct enzymes are identified.

By comparison to class II diTPSs, functional annotation of class I TPSs is inherently more complex, due to the larger size and functional diversity of the class I TPS family spanning diterpenoid as well as mono- and sesqui-terpenoid-producing enzymes. However, numerous structure-guided functional studies have provided a deeper understanding of active site determinants that control the fate of intermediary carbocations derived from ionization of the respective linear or bicyclic prenyl diphosphate substrates. Mutational studies of the *ent*-CPP/*ent*-kaurene synthases from *P. patens* and the liverwort *Jungermannia subulata* that produce *ent*-kaurene and 16 α -hydroxy-*ent*-kaurane identified an alanine residue that, when substituted for methionine or phenylalanine, blocked formation of 16 α -hydroxy-*ent*-kaurane in favor of *ent*-kaurene (Kawaide et al., 2011). Similar studies on *ent*-kaurene synthases from spruce (*Picea glauca*) and polar (*P. trichocarpa*) showed that mutagenesis of the corresponding methionine residues in these enzymes had the reciprocal effect by redirecting product specificity toward 16 α -hydroxy-*ent*-kaurane instead of *ent*-kaurene (Zerbe et al., 2012a; Irmisch

et al., 2015) (**Figure 6C**), thus suggesting a possible contribution of mutations at this position to the evolution of dedicated *ent*-kaurene synthases.

Another key active site segment that impacts class I TPS product specificity is a small hinge region between helix G1/2 (**Figure 6D**). This helix break is already present in ancestral squalene synthases, and recent structural studies of a bacterial hedycaryl sesqui-TPSs illustrated a role of this helix break in generating a negative electrostatic potential that contributes to carbocation stabilization during catalysis (Pandit et al., 2000; Baer et al., 2014). For example, mutational analysis of a pair of paralogous class II/I diTPSs from Norway spruce (*Picea abies*) illustrated that reciprocal exchange of a largely conserved SIAL/SISL motif located at this hinge region resulted in the complete interconversion of the respective abietadiene and isopimaradiene synthase activities (**Figure 6D**) (Keeling et al., 2010). A similar scenario was observed in several *ent*-kaurene synthases, where mutagenesis of a conserved isoleucine residue at this helix break mitigates formation of a tetracyclic kaurane structure and instead yielded tricyclic *ent*-pimaradiene scaffolds as demonstrated in enzymes from rice, *Arabidopsis*, spruce, and *P. patens* (Wilderman and Peters, 2007; Xu et al., 2007b; Zerbe et al., 2012a). Reciprocal mutagenesis of a corresponding threonine residue in the specialized rice class I diTPS OsKSL5 shifted catalysis from forming *ent*-pimaradiene to producing *ent*-isokaurene and other tetracyclic scaffolds, further supporting the role of residues in this position in controlling the fate of the intermediary *ent*-pimarenyl carbocation likely through electrostatic stabilization by a coordinated water or hydroxyl group (Jia et al., 2017).

More recently, mutational analysis of sclareol synthase from *S. sclarea* (Caniard et al., 2012; Schalk et al., 2012) identified a single asparagine residue, Asn431, located at the helix G break that impacts stereochemical control of product outcome. Switching Asn431 to glutamine reprogrammed the hydroxylation at C-13 from forming the native product 13*R*-sclareol to selectively producing its stereoisomer 13*S*-sclareol, thus highlighting the critical role of this amino acid on stereospecific water addition (Jia et al., 2018a). Collectively, these studies support a possibly critical role of the helix G hinge region in the catalytic control of

product specificity in distinct diTPS and likely other class I TPS enzymes from ancestral *ent*-kaurene synthases.

Numerous studies also illuminated enzyme-specific active site residues with impact on class I catalysis in mono- and sesqui-TPSs. For example, alanine substitution of Asn338 in the *Salvia fruticosa* 1,8-cineole synthase generated an enlarged active site contour and redirected catalysis to effective conversion of the C₁₅ FPP substrate to yield α -bergamotene, β -farnesene, and related sesquiterpenoid products, highlighting how minor alterations in the active cavity can enable the accommodation of different chain length substrates (Kampranis et al., 2007). In addition, a triad of active site residues impacting TPS capacity for generating cyclic products was discovered using comparative studies of *Artemisia annua* amorphadiene synthase and β -farnesene synthase that produce contrasting cyclic and linear products, respectively (Salmon et al., 2015). Large-scale site-directed mutagenesis studies of active site residues distinct between both enzymes revealed two central residue switches that activate cyclization in β -farnesene synthase (Tyr402Leu) or revert cyclization in the Tyr402Leu mutant (Val476Gly) (**Figure 6E**). Interestingly, a third mutation (Tyr430Ala) restored cyclization activity in the Val476Gly mutant background, illustrating that the ability to form a cyclic product is controlled by combinatorial effects of these active site positions. A growing body of knowledge exists on active site residues that contribute to different rearrangements of the initial cyclic carbocation intermediates in mono- and sesqui-terpenoid biosynthesis. For example, structure-guided mutagenesis of *Mentha spicata* limonene synthase, the key enzyme in the menthol production (Lange et al., 2011), identified two amino acids, His579 and Trp324, substitution of which led to premature neutralization of the carbocation intermediate to form both linear and cyclic monoterpenoids, including myrcene, linalool, and terpineol (Srividya et al., 2015) (**Figure 6F**). Similarly, reciprocal mutagenesis analyses of the mono-TPSs, Sitka spruce (*Picea sitchensis*) 3-carene synthase, and sabinene synthase associated with tree resistance against white pine weevil (Hall et al., 2011) revealed that two corresponding residues, 3-carene synthase Leu596, and sabinene synthase Phe596 located near the helix G break are critical for rearranging the central α -terpinyl⁺ carbocation toward 3-carene and sabinene, respectively (Roach et al., 2014).

The above examples and numerous related structure-function studies not covered within the scope of this review provide a mere glimpse into the plasticity of TPS catalysis, which relies on a largely non-polar active site with various possible carbocation rearrangements that enable the formation of myriad terpenoid structures with minimal investment in evolving new enzymes. However, despite these advances, our understanding of TPS mechanisms remains incomplete, thus limiting the ability to apply such knowledge for predicting the complex carbocation cascades underlying TPS activity and engineering desired enzyme functions. For instance, the taxonomic rather than functional relatedness of plant TPSs limits the use of phylogenetic analyses for functional prediction (Chen et al., 2011; Zerbe and Bohlmann, 2015). Moreover, product re-direction through TPS mutagenesis as discussed above can be accompanied by a decrease in overall enzyme activity or additional byproducts resulting from a loss

of steric control in the active site (Peters and Croteau, 2002; Pelot et al., 2016; Mafu et al., 2017). In this context, combining TPS structural analysis with quantum chemical calculations and molecular dynamic modeling approaches is advancing as a powerful tool kit to examine and predict TPS-mediated reaction cascades as discussed in more detail in several recent expert reviews (Tantillo, 2010; Tantillo, 2011; Gao et al., 2012; Major et al., 2014). For example, computational quantum chemical analyses have provided deeper insight into the inherent energy states driving terpene carbocation rearrangements and offer tools for predicting terpene pathways, as shown—for example, for predicting the often multi-product reactions catalyzed by sesquiterpene synthases (Isegawa et al., 2014). Likewise, structural studies combined with molecular dynamic modeling of TPSs has been successfully employed to predict the chemical space of possible carbocation rearrangements in mono-, di-, and tri-TPSs (Tian et al., 2014; Tian et al., 2016; Driller et al., 2018). Specifically, modeling of the catalytically relevant closed conformer of taxadiene synthase enabled important insight into the yet incompletely understood conformational changes of class I TPSs that contribute to the enzymes' control over product outcome (Schrepfer et al., 2016). In addition, detailed insights into how individual active site residues impact taxadiene synthase catalysis was revealed using a combined quantum mechanics and free energy simulation approach (Ansbacher et al., 2018). Current challenges for such computational approaches, such as predicting the role of water in the active site and the termination of the carbocation *via* deprotonation or water capture require further attention, but can likely be addressed with increasing computing resources and available structural information on a broader range of TPSs.

FUNCTIONAL ELABORATION OF THE TERPENE SCAFFOLD

The vast majority of terpenoids feature multiple functional decorations of the TPS-derived hydrocarbon scaffold that critically contribute to the diverse bioactivities of the metabolite class (Pateraki et al., 2015; Bathe and Tissier, 2019). These tailoring reactions almost invariably are initiated by position-specific oxygenations. Although these reactions can be facilitated by TPSs as outlined above, the vast majority of terpene functional modifications are controlled by the large family of cytochrome P450 monooxygenases that function as versatile catalysts for a variety of monooxygenation reactions, as well as phenol-coupling reactions, oxidative rearrangements, and oxidative C–C bond cleavage in some cases (Mizutani and Sato, 2011; Banerjee and Hamberger, 2018). Given the vast diversity of P450-controlled metabolic bifurcations, their roles in terpenoid metabolism will be on briefly discussed here. For a more expansive overview, we refer the reader to a selection of expert reviews (Nelson and Werck-Reichhart, 2011; Pateraki et al., 2015; Banerjee and Hamberger, 2018; Bathe and Tissier, 2019).

The P450 superfamily has expanded far beyond the mid-sized TPS families observed in most plants studied thus far and comprises on average more than 200 genes in an individual plant

genome with various functions in both general and specialized metabolisms. Among the 127 currently defined plant P450 families, only a handful have been shown to play major roles in terpenoid metabolism. Within the CYP85 clan, members of the CYP88A subfamily serve as *ent*-kaurenoic acid oxidases in GA biosynthesis (Nelson and Werck-Reichhart, 2011), whereas CYP725 and CYP720B enzymes are specific to gymnosperm species and catalyze hydroxylation and carboxylation reactions in the formation of taxol in species of *Taxus* and DRAs in *Pinaceae* species, respectively (Ro et al., 2005; Ro and Bohlmann, 2006; Rontein et al., 2008; Hamberger et al., 2011; Guerra-Bubb et al., 2012). More prominently, multiple families within the large CYP71 clan contribute to the various functional modifications of C₁₀–C₂₀ terpenoids (Hamberger and Bak, 2013). This includes members of the CYP701A subfamily that act as *ent*-kaurene oxidases in GA metabolism and, in several species, have been recruited through gene duplication and neo-functionalization for the formation of defensive specialized diterpenoids as exemplified in *Arabidopsis*, maize, and rice (Morrone et al., 2010; Wang et al., 2012b; Mafu et al., 2018). However, the majority of terpenoid-modifying P450s fall into the vast CYP71 and CYP76 families with numerous such enzymes having been characterized. Both P450 families are presumably evolutionary younger with the CYP76 family first occurring in cycads and *Ginkgo*, whereas the CYP71 family seemingly emerged with the onset of angiosperm evolution but is absent in nonseed plants (Nelson and Werck-Reichhart, 2011). Members of both families predominantly function as position-specific hydroxylases that catalyze (poly-)oxygenations of various mono-, sesqui-, and di-terpenoid scaffolds (Swaminathan et al., 2009; Ikezawa et al., 2011; Wu et al., 2011; Wang et al., 2012a; Diaz-Chavez et al., 2013; Guo et al., 2013; Ignea et al., 2016; Mao et al., 2016; Scheler et al., 2016; Mafu et al., 2018) but also alternate functions such as diterpenoid epoxidation and the formation of furan rings in mono- and di-terpenoid metabolisms have been demonstrated (Bertea et al., 2001; Heskes et al., 2018; Mafu et al., 2018). Notably, the first three-dimensional structure for a membrane-bound plant P450 (*S. miltiorrhiza* CYP76AH1) has been reported (Gu et al., 2019), providing resources to gain deeper mechanistic insight into the activity of diterpenoid-metabolic P450. In addition, recent P450 characterization studies expanded terpenoid-metabolic functions to other P450 families such as the gymnosperm-specific CYP750 family with a (+)-sabinene-3-oxidase (CYP750B1) from Western red cedar potentially involved in producing the anti-herbivory monoterpenoid thujone (Gesell et al., 2015), as well as members of the CYP726A subfamily from castor bean (*Ricinus communis*) that catalyze epoxidation and oxidation reactions converting macrocyclic casbene and neocembrene scaffolds in *Euphorbiaceae* species (King et al., 2014; Luo et al., 2016).

In addition to and often subsequent to the activity of P450s in the functional elaboration of terpene scaffolds, several other enzyme families contribute to the biosynthesis of bioactive terpenoids. This includes, but is not limited to, the function of 2-oxoglutarate/Fe(II)-dependent dioxygenases (2-ODDs) (Farrow and Facchini, 2014), for example, in GA phytohormone

metabolism, as well as members of often large methyl-, glycosyl-, and acetyl-transferases (Bathe and Tissier, 2019).

CONCLUDING REMARKS

Continued investigation of the evolutionary divergence and function of the TPS family will provide important knowledge of the still incompletely understood roles of terpenoids in mediating defensive and cooperative interactions with other organisms and the environment at large (Tholl, 2015). However, to address knowledge gaps and experimental limitations, research in several areas will be particularly important. Advances in the computational annotation and biochemical characterization of TPSs and P450 enzymes must continue in order to fully capitalize on rapidly expanding sequence resources across a broad range of reference and non-model species. Here, application of combinatorial functional studies in both microbial and plant hosts systems have proven to be a powerful tool to analyze modular terpenoid-metabolic networks comprised of multiple functionally distinct enzymes (Zerbe et al., 2013; Kitaoka et al., 2015; Andersen-Ranberg et al., 2016; Johnson et al., 2019a). Along with more efficient identification of new enzyme functions, continued structure-function studies will provide a deeper understanding of the functional diversity and molecular evolution of species-specific enzymes and pathways. Likewise, advanced quantum and molecular mechanics approaches for protein modeling and carbocation docking can utilize deeper structural insight to improve the precision of TPS functional prediction (Isegawa et al., 2014; Chow et al., 2015; Tian et al., 2016; O'Brien et al., 2018). Knowledge of terpenoid-metabolic genes, enzymes, and pathways will increasingly enable the investigation of terpenoid physiological functions *in planta* and under various environmental conditions. To this end, gene editing and transformation techniques applicable to a broader range of model and non-model species that produce species-specific blends of bioactive terpenoids will be critical (Wurtzel and Kutchan, 2016). Together, advanced genomic and biochemical tools and a deeper understanding of terpenoid biosynthesis and function have tremendous potential for harnessing the natural diversity of plant terpenoids for, for example, improving crop resistance and other quality traits and developing advanced protein and pathway engineering strategies for producing known and novel bioproducts.

AUTHOR CONTRIBUTIONS

PK and PZ jointly wrote the manuscript.

FUNDING

Work in the laboratory of the authors has been funded by the NSF Plant-Biotic Interactions Program (grant# 1758976 to PZ), by the DOE Joint Genome Institute Community Science Program (grant# CSP2568 to PZ), and by the DOE Early Career Research Program (grant# DE-SC0019178 to PZ).

REFERENCES

- Agrawal, A. A., and Heil, M. (2012). Synthesizing specificity: multiple approaches to understanding the attack and defense of plants. *Trends Plant Sci.* 17, 239–242. doi: 10.1016/j.tplants.2012.03.011
- Akhtar, T. A., Matsuba, Y., Schauvinhold, I., Yu, G., Lees, H. A., Klein, S. E., et al. (2013). The tomato *cis*-prenyltransferase gene family. *Plant J.* 73, 640–652. doi: 10.1111/tpj.12063
- Anarat-Cappillino, G., and Sattely, E. S. (2014). The chemical logic of plant natural product biosynthesis. *Curr. Opin. Plant Biol.* 19, 51–58. doi: 10.1016/j.pbi.2014.03.007
- Andersen-Ranberg, J., Kongstad, K. T., Nielsen, M. T., Jensen, N. B., Pateraki, I., Bach, S. S., et al. (2016). Expanding the landscape of diterpene structural diversity through stereochemically controlled combinatorial biosynthesis. *Angew. Chem. Int. Ed. Engl.* 55, 2142–2146. doi: 10.1002/anie.201510650
- Ansbacher, T., Freud, Y., and Major, D. T. (2018). Slow-starter enzymes: role of active-site architecture in the catalytic control of the biosynthesis of taxadiene by taxadiene synthase. *Biochemistry* 57, 3773–3779. doi: 10.1021/acs.biochem.8b00452
- Asawatreratanakul, K., Zhang, Y. W., Wititsuwannakul, D., Wititsuwannakul, R., Takahashi, S., Rattanapittayaporn, A., et al. (2003). Molecular cloning, expression and characterization of cDNA encoding *cis*-prenyltransferases from *Hevea brasiliensis* - A key factor participating in natural rubber biosynthesis. *Eur. J. Biochem.* 270, 4671–4680. doi: 10.1046/j.1432-1033.2003.03863.x
- Aya, K., Hiwatashi, Y., Kojima, M., Sakakibara, H., Ueguchi-Tanaka, M., Hasebe, M., et al. (2011). The Gibberellin perception system evolved to regulate a pre-existing GAMYB-mediated system during land plant evolution. *Nature Comm.* 2, 544. doi: 10.1038/ncomms1552
- Baer, P., Rabe, P., Citron, C. A., De Oliveira Mann, C. C., Kaufmann, N., Groll, M., et al. (2014). Hedycaryl synthase in complex with nerolidol reveals terpene cyclase mechanism. *ChemBiochem* 15, 213–216. doi: 10.1002/cbic.201300708
- Banerjee, A., and Hamberger, B. (2018). P450s controlling metabolic bifurcations in plant terpene specialized metabolism. *Phytochem. Rev.* 17, 81–111. doi: 10.1007/s11101-017-9530-4
- Bathe, U., and Tissier, A. (2019). Cytochrome P450 enzymes: a driving force of plant diterpene diversity. *Phytochemistry* 161, 149–162. doi: 10.1016/j.phytochem.2018.12.003
- Beran, F., Rahfeld, P., Luck, K., Nagel, R., Vogel, H., Wielsch, N., et al. (2016). Novel family of terpene synthases evolved from *trans*-isoprenyl diphosphate synthases in a flea beetle. *Proc. Natl. Acad. Sci. U. S. A.* 113, 2922–2927. doi: 10.1073/pnas.1523468113
- Berte, C. M., Schalk, M., Karp, F., Maffei, M., and Croteau, R. (2001). Demonstration that menthofuran synthase of mint (*Mentha*) is a cytochrome P450 monooxygenase: cloning, functional expression, and characterization of the responsible gene. *Arch. Biochem. Biophys.* 390, 279–286. doi: 10.1006/abbi.2001.2378
- Bessman, M. J., Frick, D. N., and O'handley, S. F. (1996). The MutT proteins or “Nudix” hydrolases, a family of versatile, widely distributed, “housecleaning” enzymes. *J. Biol. Chem.* 271, 25059–25062. doi: 10.1074/jbc.271.41.25059
- Bick, J. A., and Lange, B. M. (2003). Metabolic cross talk between cytosolic and plastidial pathways of isoprenoid biosynthesis: unidirectional transport of intermediates across the chloroplast envelope membrane. *Arch. Biochem. Biophys.* 415, 146–154. doi: 10.1016/S0003-9861(03)00233-9
- Bohlmann, J. (2011). Terpenoid synthases-from chemical ecology and forest fires to biofuels and bioproducts. *Structure* 19, 1730–1731. doi: 10.1016/j.str.2011.11.009
- Bohlmann, J., and Keeling, C. I. (2008). Terpenoid biomaterials. *Plant J.* 54, 656–669. doi: 10.1111/j.1365-313X.2008.03449.x
- Bomke, C., and Tudzynski, B. (2009). Diversity, regulation, and evolution of the gibberellin biosynthetic pathway in fungi compared to plants and bacteria. *Phytochemistry* 70, 1876–1893. doi: 10.1016/j.phytochem.2009.05.020
- Booth, J. K., and Bohlmann, J. (2019). Terpenes in *Cannabis sativa*—from plant genome to humans. *Plant Sci.* 284, 67–72. doi: 10.1016/j.plantsci.2019.03.022
- Božić, D., Papaefthimiou, D., Brückner, K., De Vos, R. C., Tsoleridis, C. A., Katsarou, D., et al. (2015). Towards elucidating carnosic acid biosynthesis in Lamiaceae: functional characterization of the three first steps of the pathway in *Salvia fruticosa* and *Rosmarinus officinalis*. *PLoS One* 10, e0124106. doi: 10.1371/journal.pone.0124106
- Brückner, K., Božić, D., Manzano, D., Papaefthimiou, D., Pateraki, I., Scheler, U., et al. (2014). Characterization of two genes for the biosynthesis of abietane-type diterpenes in rosemary (*Rosmarinus officinalis*) glandular trichomes. *Phytochemistry* 101, 52–64. doi: 10.1016/j.phytochem.2014.01.021
- Burke, C., Wildung, M. R., and Croteau, R. (1999). Geranyl diphosphate synthase: cloning, expression, and characterization of this prenyltransferase as a heterodimer. *Proc. Natl. Acad. Sci. U. S. A.* 96, 13062–13067. doi: 10.1073/pnas.96.23.13062
- Burke, C., and Croteau, R. (2002). Interaction with the small subunit of geranyl diphosphate synthase modifies the chain length specificity of geranylgeranyl diphosphate synthase to produce geranyl diphosphate. *J. Biol. Chem.* 277, 3141–3149. doi: 10.1074/jbc.M105900200
- Caniard, A., Zerbe, P., Legrand, S., Cohade, A., Valot, N., Magnard, J. L., et al. (2012). Discovery and functional characterization of two diterpene synthases for sclareol biosynthesis in *Salvia sclarea* (L.) and their relevance for perfume manufacture. *BMC Plant Biol.* 12, 119. doi: 10.1186/1471-2229-12-119
- Cao, R., Zhang, Y., Mann, F. M., Huang, C., Mukkamala, D., Hudock, M. P., et al. (2010). Diterpene cyclases and the nature of the isoprene fold. *Proteins* 78, 2417–2432. doi: 10.1002/prot.22751
- Celedon, J. M., and Bohlmann, J. (2016). Genomics-based discovery of plant genes for synthetic biology of terpenoid fragrances: a case study in sandalwood oil biosynthesis. *Methods Enzymol.* 576, 47–67. doi: 10.1016/bs.mie.2016.03.008
- Chakrabarty, R., Qu, Y., and Ro, D. K. (2015). Silencing the lettuce homologs of small rubber particle protein does not influence natural rubber biosynthesis in lettuce (*Lactuca sativa*). *Phytochemistry* 113, 121–129. doi: 10.1016/j.phytochem.2014.12.003
- Chen, F., Tholl, D., Bohlmann, J., and Pichersky, E. (2011). The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *Plant J.* 66, 212–229. doi: 10.1111/j.1365-313X.2011.04520.x
- Chen, X., Berim, A., Dayan, F. E., and Gang, D. R. (2017). A (-)-kolavenyl diphosphate synthase catalyzes the first step of salvinorin A biosynthesis in *Salvia divinorum*. *J. Exp. Bot.* 68, 1109–1122. doi: 10.1093/jxb/erw493
- Chen, X., Köllner, T. G., Jia, Q., Norris, A., Santhanam, B., Rabe, P., et al. (2016). Terpene synthase genes in eukaryotes beyond plants and fungi: occurrence in social amoebae. *Proc. Natl. Acad. Sci. U. S. A.* 113, 12132–12137. doi: 10.1073/pnas.1610379113
- Chen, X., Luck, K., Rabe, P., Dinh, C. Q., Shaulsky, G., Nelson, D. R., et al. (2019). A terpene synthase-cytochrome P450 cluster in *Dictyostelium discoideum* produces a novel trisnorsesquiterpene. *Elife* 8, e44352. doi: 10.7554/eLife.44352
- Chow, J. Y., Tian, B. X., Ramamoorthy, G., Hillerich, B. S., Seidel, R. D., Almo, S. C., et al. (2015). Computational-guided discovery and characterization of a sesquiterpene synthase from *Streptomyces clavuligerus*. *Proc. Natl. Acad. Sci. U. S. A.* 112, 5661–5666. doi: 10.1073/pnas.1505127112
- Christianson, D. W. (2008). Unearthing the roots of the terpenome. *Curr. Opin. Chem. Biol.* 12, 141–150. doi: 10.1016/j.cbpa.2007.12.008
- Christianson, D. W. (2018). Correction to structural and chemical biology of terpenoid cyclases. *Chem. Rev.* 118, 11795. doi: 10.1021/acs.chemrev.8b00682
- Cornish, K., and Xie, W. (2012). Natural rubber biosynthesis in plants: rubber transferase. *Methods Enzymol.* 515, 63–82. doi: 10.1016/B978-0-12-394290-6.00004-5
- Criswell, J., Potter, K., Shephard, F., Beale, M. H., and Peters, R. J. (2012). A single residue change leads to a hydroxylated product from the class II diterpene cyclization catalyzed by abietadiene synthase. *Org. Lett.* 14, 5828–5831. doi: 10.1021/ol3026022
- Cui, G., Duan, L., Jin, B., Qian, J., Xue, Z., Shen, G., et al. (2015). Functional divergence of diterpene synthases in the medicinal plant *Salvia miltiorrhiza*. *Plant Physiol.* 169, 1607–1618. doi: 10.1104/pp.15.00695
- D'Espaux, L., Mendez-Perez, D., Li, R., and Keasling, J. D. (2015). Synthetic biology for microbial production of lipid-based biofuels. *Curr. Opin. Chem. Biol.* 29, 58–65. doi: 10.1016/j.cbpa.2015.09.009
- Davis, E. M., and Croteau, R. (2000). “Cyclization enzymes in the biosynthesis of monoterpenes, sesquiterpenes, and diterpenes,” in *Biosynthesis. Topics Current Chem.* vol. 209. Eds. F. J. Leeper and J. C. Vederas (Berlin, Heidelberg: Springer), 53–95. doi: 10.1007/3-540-48146-X_2
- Dellas, N., Thomas, S. T., Manning, G., and Noel, J. P. (2013). Discovery of a metabolic alternative to the classical mevalonate pathway. *Elife* 2, e00672. doi: 10.7554/eLife.00672

- Diaz-Chavez, M. L., Moniodis, J., Madilao, L. L., Jancsik, S., Keeling, C. I., Barbour, E. L., et al. (2013). Biosynthesis of sandalwood oil: *Santalum album* CYP76F cytochromes P450 produce santalols and bergamotol. *PLoS One* 8, e75053. doi: 10.1371/journal.pone.0075053
- Driller, R., Janke, S., Fuchs, M., Warner, E., Mhashal, A. R., Major, D. T., et al. (2018). Towards a comprehensive understanding of the structural dynamics of a bacterial diterpene synthase during catalysis. *Nat. Commun.* 9, 3971. doi: 10.1038/s41467-018-06325-8
- Dudareva, N., Andersson, S., Orlova, I., Gatto, N., Reichelt, M., Rhodes, D., et al. (2005). The nonmevalonate pathway supports both monoterpene and sesquiterpene formation in snapdragon flowers. *Proc. Natl. Acad. Sci. U. S. A.* 102, 933–938. doi: 10.1073/pnas.0407360102
- Dudareva, N., and Pichersky, E. (2000). Biochemical and molecular genetic aspects of floral scents. *Plant Physiol.* 122, 627–633. doi: 10.1104/pp.122.3.627
- Ennajdaoui, H., Vachon, G., Giacalone, C., Besse, I., Sallaud, C., Herzog, M., et al. (2010). Trichome specific expression of the tobacco (*Nicotiana sylvestris*) cembratrien-ol synthase genes is controlled by both activating and repressing cis-regions. *Plant Mol. Biol.* 73, 673–685. doi: 10.1007/s11103-010-9648-x
- Falara, V., Alba, J. M., Kant, M. R., Schuurink, R. C., and Pichersky, E. (2014). Geranylinalool synthases in Solanaceae and other angiosperms constitute an ancient branch of diterpene synthases involved in the synthesis of defensive compounds. *Plant Physiol.* 166, 428–441. doi: 10.1104/pp.114.243246
- Falara, V., Pichersky, E., and Kanellis, A. K. (2010). A copal-8-ol diphosphate synthase from the angiosperm *Cistus creticus* subsp. *creticus* is a putative key enzyme for the formation of pharmacologically active, oxygen-containing labdane-type diterpenes. *Plant Physiol.* 154, 301–310. doi: 10.1104/pp.110.159566
- Farrow, S. C., and Fachini, P. J. (2014). Functional diversity of 2-oxoglutarate/Fe(II)-dependent dioxygenases in plant metabolism. *Front. Plant Sci.* 5, 524. doi: 10.3389/fpls.2014.00524
- Fischer, M. J., Rustenhloz, C., Leh-Louis, V., and Perriere, G. (2015). Molecular and functional evolution of the fungal diterpene synthase genes. *BMC Microbiol.* 15, 221. doi: 10.1186/s12866-015-0564-8
- Flugge, U. I., and Gao, W. (2005). Transport of isoprenoid intermediates across chloroplast envelope membranes. *Plant Biol.* 7, 91–97. doi: 10.1055/s-2004-830446
- Fu, J., Ren, F., Lu, X., Mao, H., Xu, M., and Degenhardt, J. (2016). A tandem array of *ent*-kaurene synthases in maize with roles in gibberellin and more specialized metabolism. *Plant Physiol.* 170, 742–751. doi: 10.1104/pp.15.01727
- Gao, W., Sun, H. X., Xiao, H., Cui, G., Hillwig, M. L., Jackson, A., et al. (2014). Combining metabolomics and transcriptomics to characterize tanshinone biosynthesis in *Salvia miltiorrhiza*. *BMC Genomics* 15, 73. doi: 10.1186/1471-2164-15-73
- Gao, Y., Honzatko, R. B., and Peters, R. J. (2012). Terpenoid synthase structures: a so far incomplete view of complex catalysis. *Nat. Prod. Rep.* 29, 1153–1175. doi: 10.1039/c2np20059g
- Gennadios, H. A., Gonzalez, V., Di Costanzo, L., Li, A., Yu, F., Miller, D. J., et al. (2009). Crystal structure of (+)-delta-cadinene synthase from *Gossypium arboreum* and evolutionary divergence of metal binding motifs for catalysis. *Biochemistry* 48, 6175–6183. doi: 10.1021/bi900483b
- Gershenzon, J., and Dudareva, N. (2007). The function of terpene natural products in the natural world. *Nat. Chem. Biol.* 3, 408–414. doi: 10.1038/nchembio.2007.5
- Gesell, A., Blaukopf, M., Madilao, L., Yuen, M. M., Withers, S. G., Mattsson, J., et al. (2015). The gymnosperm cytochrome P450 CYP750B1 catalyzes stereospecific monoterpene hydroxylation of (+)-sabinene in thujone biosynthesis in western red cedar. *Plant Physiol.* 168, 94–106. doi: 10.1104/pp.15.00315
- Ghassemlian, M., Lutes, J., Tepperman, J. M., Chang, H. S., Zhu, T., Wang, X., et al. (2006). Integrative analysis of transcript and metabolite profiling data sets to evaluate the regulation of biochemical pathways during photomorphogenesis. *Arch. Biochem. Biophys.* 448, 45–59. doi: 10.1016/j.abb.2005.11.020
- Gross, H., and König, G. M. (2006). Terpenoids from marine organisms: unique structures and their pharmacological potential. *Phytochem. Rev.* 5, 115–141. doi: 10.1007/s11101-005-5464-3
- Gu, M., Wang, M., Guo, J., Shi, C., Deng, J., Huang, L., et al. (2019). Crystal structure of CYP76AH1 in 4-PI-bound state from *Salvia miltiorrhiza*. *Biochem. Biophys. Res. Commun.* 511, 813–819. doi: 10.1016/j.bbrc.2019.02.103
- Guerra-Bubb, J., Croteau, R., and Williams, R. M. (2012). The early stages of taxol biosynthesis: an interim report on the synthesis and identification of early pathway metabolites. *Nat. Prod. Rep.* 29, 683–696. doi: 10.1039/c2np20021j
- Guo, J., Zhou, Y. J., Hillwig, M. L., Shen, Y., Yang, L., Wang, Y., et al. (2013). CYP76AH1 catalyzes turnover of miltiradiene in tanshinones biosynthesis and enables heterologous production of ferruginol in yeasts. *Proc. Natl. Acad. Sci. U. S. A.* 110, 12108–12113. doi: 10.1073/pnas.1218061110
- Gutensohn, M., Orlova, I., Nguyen, T. T., Davidovich-Rikanati, R., Ferruzzi, M. G., Sitrit, Y., et al. (2013). Cytosolic monoterpene biosynthesis is supported by plastid-generated geranyl diphosphate substrate in transgenic tomato fruits. *Plant J.* 75, 351–363. doi: 10.1111/tpj.12212
- Hall, D. E., Robert, J. A., Keeling, C. I., Domanski, D., Quesada, A. L., Jancsik, S., et al. (2011). An integrated genomic, proteomic and biochemical analysis of (+)-3-carene biosynthesis in Sitka spruce (*Picea sitchensis*) genotypes that are resistant or susceptible to white pine weevil. *Plant J.* 65, 936–948. doi: 10.1111/j.1365-3113.2010.04478.x
- Hall, D. E., Zerbe, P., Jancsik, S., Quesada, A. L., Dullat, H., Madilao, L. L., et al. (2013). Evolution of conifer diterpene synthases: diterpene resin acid biosynthesis in lodgepole pine and jack pine involves monofunctional and bifunctional diterpene synthases. *Plant Physiol.* 161, 600–616. doi: 10.1104/pp.112.208546
- Hamberger, B., and Bak, S. (2013). Plant P450s as versatile drivers for evolution of species-specific chemical diversity. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 368, 20120426. doi: 10.1098/rstb.2012.0426
- Hamberger, B., Ohnishi, T., Hamberger, B., Seguin, A., and Bohlmann, J. (2011). Evolution of diterpene metabolism: Sitka spruce CYP720B4 catalyzes multiple oxidations in resin acid biosynthesis of conifer defense against insects. *Plant Physiol.* 157, 1677–1695. doi: 10.1104/pp.111.185843
- Hampel, D., Mosandl, A., and Wust, M. (2005). Biosynthesis of mono- and sesquiterpenes in carrot roots and leaves (*Daucus carota* L.): metabolic cross talk of cytosolic mevalonate and plastidial methylerythritol phosphate pathways. *Phytochemistry* 66, 305–311. doi: 10.1016/j.phytochem.2004.12.010
- Hansen, N. L., Heskes, A. M., Hamberger, B., Olsen, C. E., Hallström, B. M., Andersen-Ranberg, J., et al. (2017a). The terpene synthase gene family in *Tripterygium wilfordii* harbors a labdane-type diterpene synthase among the monoterpene synthase TPS-b subfamily. *Plant J.* 89, 429–441. doi: 10.1111/tpj.13410
- Hansen, N. L., Nissen, J. N., and Hamberger, B. (2017b). Two residues determine the product profile of the class II diterpene synthases TPS14 and TPS21 of *Tripterygium wilfordii*. *Phytochemistry* 138, 52–56. doi: 10.1016/j.phytochem.2017.02.022
- Hayashi, K., Kawaide, H., Notomi, M., Sakigi, Y., Matsuo, A., and Nozaki, H. (2006). Identification and functional analysis of bifunctional *ent*-kaurene synthase from the moss *Physcomitrella patens*. *FEBS Lett.* 580, 6175–6181. doi: 10.1016/j.febslet.2006.10.018
- Heil, M., and Ton, J. (2008). Long-distance signalling in plant defence. *Trends Plant Sci.* 13, 264–272. doi: 10.1016/j.tplants.2008.03.005
- Helmstädter, A. (2013). *Tripterygium wilfordii* Hook. f.—how a traditional Taiwanese medicinal plant found its way to the West. *Pharmazie* 68, 643–646. doi: 10.1691/ph.2013.6506
- Hemmerlin, A., Harwood, J. L., and Bach, T. J. (2012). A raison d'être for two distinct pathways in the early steps of plant isoprenoid biosynthesis? *Prog. Lipid Res.* 51, 95–148. doi: 10.1016/j.plipres.2011.12.001
- Hemmerlin, A., Hoeffler, J. F., Meyer, O., Tritsch, D., Kagan, I. A., Grosdemange-Billiard, C., et al. (2003). Cross-talk between the cytosolic mevalonate and the plastidial methylerythritol phosphate pathways in tobacco bright yellow-2 cells. *J. Biol. Chem.* 278, 26666–26676. doi: 10.1074/jbc.M302526200
- Henry, L. K., Gutensohn, M., Thomas, S. T., Noel, J. P., and Dudareva, N. (2015). Orthologs of the archaeal isopentenyl phosphate kinase regulate terpenoid production in plants. *Proc. Natl. Acad. Sci. U. S. A.* 112, 10050–10055. doi: 10.1073/pnas.1504798112
- Henry, L. K., Thomas, S. T., Widhalm, J. R., Lynch, J. H., Davis, T. C., Kessler, S. A., et al. (2018). Contribution of isopentenyl phosphate to plant terpenoid metabolism. *Nat. Plants* 4, 721–729. doi: 10.1038/s41477-018-0220-z
- Heskes, A. M., Sundram, T. C. M., Boughton, B. A., Jensen, N. B., Hansen, N. L., Crocoll, C., et al. (2018). Biosynthesis of bioactive diterpenoids in the medicinal plant *Vitex agnus-castus*. *Plant J.* 93, 943–958. doi: 10.1111/tpj.13822

- Hillwig, M. L., Xu, M., Toyomasu, T., Tiernan, M. S., Wei, G., Cui, G., et al. (2011). Domain loss has independently occurred multiple times in plant terpenes synthase evolution. *Plant J.* 68, 1051–1060. doi: 10.1111/j.1365-3113X.2011.04756.x
- Hsieh, F. L., Chang, T. H., Ko, T. P., and Wang, A. H. (2011). Structure and mechanism of an Arabidopsis medium/long-chain-length prenyl pyrophosphate synthase. *Plant Physiol.* 155, 1079–1090. doi: 10.1104/pp.110.168799
- Huang, A. C., Kautsar, S. A., Hong, Y. J., Medema, M. H., Bond, A. D., Tantillo, D. J., et al. (2017). Unearthing a sesterterpene biosynthetic repertoire in the Brassicaceae through genome mining reveals convergent evolution. *Proc. Natl. Acad. Sci. U. S. A.* 114, E6005–E6014. doi: 10.1073/pnas.1705567114
- Hyatt, D. C., Youn, B., Zhao, Y., Santhamma, B., Coates, R. M., Croteau, R. B., et al. (2007). Structure of limonene synthase, a simple model for terpenoid cyclase catalysis. *Proc. Natl. Acad. Sci. U. S. A.* 104, 5360–5365. doi: 10.1073/pnas.0700915104
- Ignia, C., Athanasakoglou, A., Ioannou, E., Georgantea, P., Trikkas, F. A., Loupassaki, S., et al. (2016). Carnosic acid biosynthesis elucidated by a synthetic biology platform. *Proc. Natl. Acad. Sci. U. S. A.* 113, 3681–3686. doi: 10.1073/pnas.1523787113
- Ikezawa, N., Gopfert, J. C., Nguyen, D. T., Kim, S. U., O'maille, P. E., Spring, O., et al. (2011). Lettuce costunolide synthase (CYP71BL2) and its homolog (CYP71BL1) from sunflower catalyze distinct regio- and stereoselective hydroxylations in sesquiterpene lactone metabolism. *J. Biol. Chem.* 286, 21601–21611. doi: 10.1074/jbc.M110.216804
- Irmisch, S., Muller, A. T., Schmidt, L., Gunther, J., Gershenzon, J., and Köllner, T. G. (2015). One amino acid makes the difference: the formation of *ent*-kaurene and 16 α -hydroxy-*ent*-kaurene by diterpene synthases in poplar. *BMC Plant Biol.* 15, 262. doi: 10.1186/s12870-015-0647-6
- Isegawa, M., Maeda, S., Tantillo, D. J., and Morokuma, K. (2014). Predicting pathways for terpene formation from first principles—routes to known and new sesquiterpenes. *Chem. Sci.* 5, 1555–1560. doi: 10.1039/c3sc53293c
- Jia, M., O'Brien, T. E., Zhang, Y., Siegel, J. B., Tantillo, D. J., and Peters, R. J. (2018a). Changing face: a key residue for the addition of water by sclareol synthase. *ACS Catal.* 8, 3133–3137. doi: 10.1021/acscatal.8b00121
- Jia, M., and Peters, R. J. (2017). *Cis* or *trans* with class II diterpene cyclases. *Org. Biomol. Chem.* 15, 3158–3160. doi: 10.1039/C7OB00510E
- Jia, M., Potter, K. C., and Peters, R. J. (2016a). Extreme promiscuity of a bacterial and a plant diterpene synthase enables combinatorial biosynthesis. *Metab. Eng.* 37, 24–34. doi: 10.1016/j.ymben.2016.04.001
- Jia, M., Zhou, K., Tufts, S., Schulte, S., and Peters, R. J. (2017). A pair of residues that interactively affect diterpene synthase product outcome. *ACS Chem. Biol.* 12, 862–867. doi: 10.1021/acscchembio.6b01075
- Jia, Q., Köllner, T. G., Gershenzon, J., and Chen, F. (2018b). MTPSLs: new terpene synthases in nonseed plants. *Trends Plant Sci.* 23, 121–128. doi: 10.1016/j.tplants.2017.09.014
- Jia, Q., Li, G., Köllner, T. G., Fu, J., Chen, X., Xiong, W., et al. (2016b). Microbial-type terpene synthase genes occur widely in nonseed land plants, but not in seed plants. *Proc. Natl. Acad. Sci. U. S. A.* 113, 12328–12333. doi: 10.1073/pnas.1607973113
- Johnson, S. R., Bhat, W. W., Bibik, J., Turmo, A., Hamberger, B., Evolutionary Mint Genomics Consortium, et al. (2019a). A database-driven approach identifies additional diterpene synthase activities in the mint family (Lamiaceae). *J. Biol. Chem.* 294, 1349–1362. doi: 10.1074/jbc.RA118.006025
- Johnson, S. R., Bhat, W. W., Sadre, R., Miller, G. P., Garcia, A. S., and Hamberger, B. (2019b). Promiscuous terpene synthases from *Prunella vulgaris* highlight the importance of substrate and compartment switching in terpene synthase evolution. *New Phytol.* 23, 323–335. doi: 10.1111/nph.15778
- Kampranis, S. C., Ioannidis, D., Purvis, A., Mahrez, W., Ninga, E., Katerelos, N. A., et al. (2007). Rational conversion of substrate and product specificity in a *Salvia* monoterpene synthase: structural insights into the evolution of terpene synthase function. *Plant Cell* 19, 1994–2005. doi: 10.1105/tpc.106.047779
- Kato-Noguchi, H., and Peters, R. J. (2013). The role of momilactones in rice allelopathy. *J. Chem. Ecol.* 39, 175–185. doi: 10.1007/s10886-013-0236-9
- Kawaide, H., Hayashi, K., Kawanabe, R., Sakigi, Y., Matsuo, A., Natsume, M., et al. (2011). Identification of the single amino acid involved in quenching the *ent*-kauranil cation by a water molecule in *ent*-kaurene synthase of *Physcomitrella patens*. *FEBS J.* 278, 123–133. doi: 10.1111/j.1742-4658.2010.07938.x
- Keasling, J. D. (2012). Synthetic biology and the development of tools for metabolic engineering. *Metab. Eng.* 14, 189–195. doi: 10.1016/j.ymben.2012.01.004
- Keeling, C. I., and Bohlmann, J. (2006a). Diterpene resin acids in conifers. *Phytochemistry* 67, 2415–2423. doi: 10.1016/j.phytochem.2006.08.019
- Keeling, C. I., and Bohlmann, J. (2006b). Genes, enzymes and chemicals of terpenoid diversity in the constitutive and induced defence of conifers against insects and pathogens. *New Phytol.* 170, 657–675. doi: 10.1111/j.1469-8137.2006.01716.x
- Keeling, C. I., Dullat, H. K., Yuen, M., Ralph, S. G., Jancsik, S., and Bohlmann, J. (2010). Identification and functional characterization of monofunctional *ent*-CPP and *ent*-kaurene synthases in white spruce reveal different patterns for diterpene synthase evolution for primary and secondary metabolism in gymnosperms. *Plant Physiol.* 152, 1197–1208. doi: 10.1104/pp.109.151456
- Keeling, C. I., Madilao, L. L., Zerbe, P., Dullat, H. K., and Bohlmann, J. (2011). The primary diterpene synthase products of *Picea abies* levopimaradiene/abietadiene synthase (PaLAS) are epimers of a thermally unstable diterpenol. *J. Biol. Chem.* 286, 21145–21153. doi: 10.1074/jbc.M111.245951
- Keeling, C. I., Weisshaar, S., Lin, R. P., and Bohlmann, J. (2008). Functional plasticity of paralogous diterpene synthases involved in conifer defense. *Proc. Natl. Acad. Sci. U. S. A.* 105, 1085–1090. doi: 10.1073/pnas.0709466105
- Kersten, R. D., Lee, S., Fujita, D., Pluskal, T., Kram, S., Smith, J. E., et al. (2017). A red algal bourbonane sesquiterpene synthase defined by microgram-scale NMR-coupled crystalline sponge X-ray diffraction analysis. *J. Am. Chem. Soc.* 139, 16838–16844. doi: 10.1021/jacs.7b09452
- King, A. J., Brown, G. D., Gilday, A. D., Larson, T. R., and Graham, I. A. (2014). Production of bioactive diterpenoids in the Euphorbiaceae depends on evolutionarily conserved gene clusters. *Plant Cell* 26, 3286–3298. doi: 10.1105/tpc.114.129668
- Kirby, J., Nishimoto, M., Park, J. G., Withers, S. T., Nowroozi, F., Behrendt, D., et al. (2010). Cloning of casbene and neocembrene synthases from Euphorbiaceae plants and expression in *Saccharomyces cerevisiae*. *Phytochemistry* 71, 1466–1473. doi: 10.1016/j.phytochem.2010.06.001
- Kitaoka, N., Lu, X., Yang, B., and Peters, R. J. (2015). The application of synthetic biology to elucidation of plant mono-, sesqui-, and diterpenoid metabolism. *Mol. Plant* 8, 6–16. doi: 10.1016/j.molp.2014.12.002
- Köksal, M., Jin, Y., Coates, R. M., Croteau, R., and Christianson, D. W. (2011). Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis. *Nature* 469, 116–120. doi: 10.1038/nature09628
- Köksal, M., Potter, K., Peters, R. J., and Christianson, D. W. (2014). 1.55 Å-resolution structure of *ent*-copalyl diphosphate synthase and exploration of general acid function by site-directed mutagenesis. *Biochim. Biophys. Acta* 1840, 184–190. doi: 10.1016/j.bbagen.2013.09.004
- Külheim, C., Padovan, A., Hefer, C., Krause, S. T., Köllner, T. G., Myburg, A. A., et al. (2015). The Eucalyptus terpene synthase gene family. *BMC Genomics* 16, 450. doi: 10.1186/s12864-015-1598-x
- Kumar, S., Kempinski, C., Zhuang, X., Norris, A., Mafu, S., Zi, J., et al. (2016). Molecular diversity of terpene synthases in the Liverwort *Marchantia polymorpha*. *Plant Cell* 28, 2632–2650. doi: 10.1105/tpc.16.00062
- Lancaster, J., Khimian, A., Young, S., Lehner, B., Luck, K., Wallingford, A., et al. (2018). De novo formation of an aggregation pheromone precursor by an isoprenyl diphosphate synthase-related terpene synthase in the harlequin bug. *Proc. Natl. Acad. Sci. U. S. A.* 115, E8634–E8641. doi: 10.1073/pnas.1800008115
- Lange, B. M., Mahmoud, S. S., Wildung, M. R., Turner, G. W., Davis, E. M., Lange, I., et al. (2011). Improving peppermint essential oil yield and composition by metabolic engineering. *Proc. Natl. Acad. Sci. U. S. A.* 108, 16944–16949. doi: 10.1073/pnas.1111558108
- Lange, B. M., Rujan, T., Martin, W., and Croteau, R. (2000). Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. *Proc. Natl. Acad. Sci. U. S. A.* 97, 13172–13177. doi: 10.1073/pnas.240454797
- Lange, B. M., and Turner, G. W. (2013). Terpenoid biosynthesis in trichomes: current status and future opportunities. *Plant Biotechnol. J.* 11, 2–22. doi: 10.1111/j.1467-7652.2012.00737.x
- Laule, O., Furlholz, A., Chang, H. S., Zhu, T., Wang, X., Heifetz, P. B., et al. (2003). Crosstalk between cytosolic and plastidial pathways of isoprenoid biosynthesis in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.* 100, 6866–6871. doi: 10.1073/pnas.1031755100
- Li, G., Köllner, T. G., Yin, Y., Jiang, Y., Chen, H., Xu, Y., et al. (2012). Nonseed plant *Selaginella moellendorffii* has both seed plant and microbial types of terpene synthases. *Proc. Natl. Acad. Sci. U. S. A.* 109, 14711–14715. doi: 10.1073/pnas.1204300109

- Liang, P. H., Ko, T. P., and Wang, A. H. (2002). Structure, mechanism and function of prenyltransferases. *Eur. J. Biochem.* 269, 3339–3354. doi: 10.1046/j.1432-1033.2002.03014.x
- Liu, W., Feng, X., Zheng, Y., Huang, C. H., Nakano, C., Hoshino, T., et al. (2014). Structure, function and inhibition of *ent*-kaurene synthase from *Bradyrhizobium japonicum*. *Sci. Rep.* 4, 6214. doi: 10.1038/srep06214
- Lohr, M., Schwender, J., and Polle, J. E. (2012). Isoprenoid biosynthesis in eukaryotic phototrophs: a spotlight on algae. *Plant Sci.* 185–186, 9–22. doi: 10.1016/j.plantsci.2011.07.018
- Lombard, J., and Moreira, D. (2011). Origins and early evolution of the mevalonate pathway of isoprenoid biosynthesis in the three domains of life. *Mol. Biol. Evol.* 28, 87–99. doi: 10.1093/molbev/msq177
- Lopez, M. L., Bonzani, N. E., and Zygadlo, J. A. (2008). Allelopathic potential of *Tagetes minuta* terpenes by a chemical, anatomical and phytotoxic approach. *Biochem. Sys. Ecol.* 36, 882–890. doi: 10.1016/j.bse.2008.11.003
- Luo, D., Callari, R., Hamberger, B., Wubshet, S. G., Nielsen, M. T., Andersen-Ranberg, J., et al. (2016). Oxidation and cyclization of casbene in the biosynthesis of Euphorbia factors from mature seeds of *Euphorbia lathyris* L. *Proc. Natl. Acad. Sci. U. S. A.* 113, E5082–E5089. doi: 10.1073/pnas.1607504113
- Luo, S. H., Luo, Q., Niu, X. M., Xie, M. J., Zhao, X., Schneider, B., et al. (2010). Glandular trichomes of *Leucosceptum canum* harbor defensive sesterterpenoids. *Angew. Chem. Int. Ed. Engl.* 49, 4471–4475. doi: 10.1002/anie.201000449
- Ma, Y., Yuan, L., Wu, B., Li, X., Chen, S., and Lu, S. (2012). Genome-wide identification and characterization of novel genes involved in terpenoid biosynthesis in *Salvia miltiorrhiza*. *J. Exp. Bot.* 63, 2809–2823. doi: 10.1093/jxb/err466
- Mafu, S., Ding, Y., Murphy, K. M., Yaacoobi, O., Addison, J. B., Wang, Q., et al. (2018). Discovery, biosynthesis and stress-related accumulation of dolabradene-derived defenses in maize. *Plant Physiol.* 176, 2677–2690. doi: 10.1104/pp.17.01351
- Mafu, S., Fischer, E., Addisson, B., Riberio Barbosana, I., and Zerbe, P. (2016). Substitution of two active site residues alters C9-hydroxylation in a class II diterpene synthase. *Chembiochem* 17, 2304–2307. doi: 10.1002/cbic.201600419
- Mafu, S., Hillwig, M. L., and Peters, R. J. (2011). A novel labda-7,13E-dien-15-ol-producing bifunctional diterpene synthase from *Selaginella moellendorffii*. *Chembiochem* 12, 1984–1987. doi: 10.1002/cbic.201100336
- Mafu, S., Karunanithi, P. S., Palazzo, T. A., Harrod, B. L., Rodriguez, S. M., Mollhoff, I. N., et al. (2017). Biosynthesis of the microtubule-destabilizing diterpene pseudolaric acid B from golden larch involves an unusual diterpene synthase. *Proc. Natl. Acad. Sci. U. S. A.* 114, 974–979. doi: 10.1073/pnas.1612901114
- Mafu, S., Potter, K. C., Hillwig, M. L., Schulte, S., Criswell, J., and Peters, R. J. (2015). Efficient heterocyclisation by (di)terpene synthases. *Chem. Commun.* 51, 13485–13487. doi: 10.1039/C5CC05754J
- Mafu, S., and Zerbe, P. (2018). Plant diterpenoid metabolism for manufacturing the biopharmaceuticals of tomorrow: prospects and challenges. *Phytochem. Rev.* 17, 113–130. doi: 10.1007/s11101-017-9513-5
- Magnard, J. L., Rocca, A., Caissard, J. C., Vergne, P., Sun, P., Hecquet, R., et al. (2015). PLANT VOLATILES. Biosynthesis of monoterpene scent compounds in roses. *Science* 349, 81–83. doi: 10.1126/science.aab0696
- Major, D. T., Freud, Y., and Weitman, M. (2014). Catalytic control in terpenoid cyclases: multiscale modeling of thermodynamic, kinetic, and dynamic effects. *Curr. Opin. Chem. Biol.* 21, 25–33. doi: 10.1016/j.cbpa.2014.03.010
- Mann, F. M., Pristic, S., Davenport, E. K., Determan, M. K., Coates, R. M., and Peters, R. J. (2010). A single residue switch for Mg⁽²⁺⁾-dependent inhibition characterizes plant class II diterpene cyclases from primary and secondary metabolism. *J. Biol. Chem.* 285, 20558–20563. doi: 10.1074/jbc.M110.123307
- Mao, H., Liu, J., Ren, F., Peters, R. J., and Wang, Q. (2016). Characterization of CYP71Z18 indicates a role in maize zealexin biosynthesis. *Phytochemistry* 121, 4–10. doi: 10.1016/j.phytochem.2015.10.003
- Martin, D. M., Aubourg, S., Schouwey, M. B., Daviet, L., Schalk, M., Toub, O., et al. (2010). Functional annotation, genome organization and phylogeny of the grapevine (*Vitis vinifera*) terpene synthase gene family based on genome assembly, FLcDNA cloning, and enzyme assays. *BMC Plant Biol.* 10, 226. doi: 10.1186/1471-2229-10-226
- Martin, D. M., Fäldt, J., and Bohlmann, J. (2004). Functional characterization of nine Norway Spruce TPS genes and evolution of gymnosperm terpene synthases of the TPS-d subfamily. *Plant Physiol.* 135, 1908–1927. doi: 10.1104/pp.104.042028
- Matsuba, Y., Zi, J., Jones, A. D., Peters, R. J., and Pichersky, E. (2015). Biosynthesis of the diterpenoid lycosantalanol via neryleryl diphosphate in *Solanum lycopersicum*. *PLoS One* 10, e0119302. doi: 10.1371/journal.pone.0119302
- Matsumi, R., Atomi, H., Driessen, A. J. M., and Van Der Oost, J. (2011). Isoprenoid biosynthesis in Archaea—biochemical and evolutionary implications. *Res. Microbiol.* 162, 39–52. doi: 10.1016/j.resmic.2010.10.003
- Mau, C. J., and West, C. A. (1994). Cloning of casbene synthase cDNA: evidence for conserved structural features among terpenoid cyclases in plants. *Proc. Natl. Acad. Sci. U. S. A.* 91, 8497–8501. doi: 10.1073/pnas.91.18.8497
- McAndrew, R. P., Peralta-Yahya, P. P., Degiovanni, A., Pereira, J. H., Hadi, M. Z., Keasling, J. D., et al. (2011). Structure of a three-domain sesquiterpene synthase: a prospective target for advanced biofuels production. *Structure* 19, 1876–1884. doi: 10.1016/j.str.2011.09.013
- McGarvey, D. J., and Croteau, R. (1995). Terpenoid metabolism. *Plant Cell* 7, 1015–1026. doi: 10.1105/tpc.7.7.1015
- Mendoza-Poudereux, I., Kutzner, E., Huber, C., Segura, J., Eisenreich, W., and Arrillaga, I. (2015). Metabolic cross-talk between pathways of terpenoid backbone biosynthesis in spike lavender. *Plant Physiol. Biochem.* 95, 113–120. doi: 10.1016/j.plaphy.2015.07.029
- Minami, A., Ozaki, T., Liu, C., and Oikawa, H. (2018). Cyclopentane-forming di/terpene synthases: widely distributed enzymes in bacteria, fungi, and plants. *Nat. Prod. Rep.* 35, 1330–1346. doi: 10.1039/C8NP00026C
- Mizutani, M., and Sato, F. (2011). Unusual P450 reactions in plant secondary metabolism. *Arch. Biochem. Biophys.* 507, 194–203. doi: 10.1016/j.abb.2010.09.026
- Mitsuhashi, T., and Abe, I. (2018). Chimeric terpene synthases possessing both terpene cyclization and prenyltransfer activities. *Chembiochem* 110601114. doi: 10.1002/cbic.201800120
- Morrone, D., Chambers, J., Lowry, L., Kim, G., Anterola, A., Bender, K., et al. (2009). Gibberellin biosynthesis in bacteria: separate *ent*-copalyl diphosphate and *ent*-kaurene synthases in *Bradyrhizobium japonicum*. *FEBS Lett.* 583, 475–480. doi: 10.1016/j.febslet.2008.12.052
- Morrone, D., Chen, X., Coates, R. M., and Peters, R. J. (2010). Characterization of the kaurene oxidase CYP701A3, a multifunctional cytochrome P450 from gibberellin biosynthesis. *Biochem. J.* 431, 337–344. doi: 10.1042/BJ20100597
- Morrone, D., Hillwig, M. L., Mead, M. E., Lowry, L., Fulton, D. B., and Peters, R. J. (2011). Evident and latent plasticity across the rice diterpene synthase family with potential implications for the evolution of diterpenoid metabolism in the cereals. *Biochem. J.* 435, 589–595. doi: 10.1042/BJ20101429
- Murphy, K. M., Ma, L. T., Ding, Y., Schmelz, E. A., and Zerbe, P. (2018). Functional characterization of two class II diterpene synthases indicates additional specialized diterpenoid pathways in maize (*Zea mays*). *Front. Plant Sci.* 9, 1542. doi: 10.3389/fpls.2018.01542
- Nagel, R., Bernholz, C., Vranova, E., Kosuth, J., Bergau, N., Ludwig, S., et al. (2015). *Arabidopsis thaliana* isoprenyl diphosphate synthases produce the C25 intermediate geranylarnesyl diphosphate. *Plant J.* 84, 847–859. doi: 10.1111/tpj.13064
- Nagel, R., Bieber, J. E., Schmidt-Dannert, M. G., Nett, R. S., and Peters, R. J. (2018). A third class: functional gibberellin biosynthetic operon in beta-proteobacteria. *Front. Microbiol.* 9, 2916. doi: 10.3389/fmicb.2018.02916
- Nagel, R., and Peters, R. J. (2017). Investigating the Phylogenetic range of gibberellin biosynthesis in bacteria. *Mol. Plant Microbe Interact.* 30, 343–349. doi: 10.1094/MPMI-01-17-0001-R
- Nagel, R., Schmidt, A., and Peters, R. J. (2019). Isoprenyl diphosphate synthases: the chain length determining step in terpene biosynthesis. *Planta* 249, 9–20. doi: 10.1007/s00425-018-3052-1
- Nelson, D., and Werck-Reichhart, D. (2011). A P450-centric view of plant evolution. *Plant J.* 66, 194–211. doi: 10.1111/j.1365-3113.2011.04529.x
- O'Brien, T. E., Bertolani, S. J., Zhang, Y., Siegel, J. B., and Tantillo, D. J. (2018). Predicting productive binding modes for substrates and carbocation intermediates in terpene synthases-bornyl diphosphate synthase as a representative case. *ACS Catal.* 8, 3322–3330. doi: 10.1021/acscatal.8b00342
- Ogawa, T., Ueda, Y., Yoshimura, K., and Shigeoka, S. (2005). Comprehensive analysis of cytosolic Nudix hydrolases in *Arabidopsis thaliana*. *J. Biol. Chem.* 280, 25277–25283. doi: 10.1074/jbc.M503536200

- Oh, S. K., Han, K. H., Ryu, S. B., and Kang, H. (2000). Molecular cloning, expression, and functional analysis of a *cis*-prenyltransferase from *Arabidopsis thaliana*—implications in rubber biosynthesis. *J. Biol. Chem.* 275, 18482–18488. doi: 10.1074/jbc.M002000200
- Okada, K., Kawaide, H., Miyamoto, K., Miyazaki, S., Kainuma, R., Kimura, H., et al. (2016). HpDTC1, a stress-inducible bifunctional diterpene cyclase involved in momilactone biosynthesis, functions in chemical defence in the moss *Hypnum plumaeforme*. *Sci. Rep.* 6, 25316. doi: 10.1038/srep25316
- Opitz, S., Nes, W. D., and Gershenzon, J. (2014). Both methylerythritol phosphate and mevalonate pathways contribute to biosynthesis of each of the major isoprenoid classes in young cotton seedlings. *Phytochemistry* 98, 110–119. doi: 10.1016/j.phytochem.2013.11.010
- Otomo, K., Kenmoku, H., Oikawa, H., Konig, W. A., Toshima, H., Mitsuhashi, W., et al. (2004). Biological functions of *ent*- and *syn*-copalyl diphosphate synthases in rice: key enzymes for the branch point of gibberellin and phytoalexin biosynthesis. *Plant J.* 39, 886–893. doi: 10.1111/j.1365-313X.2004.02175.x
- Ourisson, G., and Nakatani, Y. (1994). The terpenoid theory of the origin of cellular life: the evolution of terpenoids to cholesterol. *Chem. Biol.* 1, 11–23. doi: 10.1016/1074-5521(94)90036-1
- Paddon, C. J., Westfall, P. J., Pitera, D. J., Benjamin, K., Fisher, K., McPhee, D., et al. (2013). High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* 496, 528–532. doi: 10.1038/nature12051
- Pallardy, S. G. (2008). “Plant hormones and other signaling molecules,” in *Physiology of woody plants*, 3rd edition. Editor S. G. Pallardy (Columbia, Missouri: Academic Press), 367–377. doi: 10.1016/B978-012088765-1.50014-2
- Pandit, J., Danley, D. E., Schulte, G. K., Mazzalupo, S., Pauly, T. A., Hayward, C. M., et al. (2000). Crystal structure of human squalene synthase. A key enzyme in cholesterol biosynthesis. *J. Biol. Chem.* 275, 30610–30617. doi: 10.1074/jbc.M004132200
- Pateraki, I., Andersen-Ranberg, J., Jensen, N. B., Wubshet, S. G., Heskes, A. M., Forman, V., et al. (2017). Total biosynthesis of the cyclic AMP booster forskolin from *Coleus forskohlii*. *Elife* 6, e23001. doi: 10.7554/eLife.23001
- Pateraki, I., Heskes, A. M., and Hamberger, B. (2015). Cytochromes P450 for terpene functionalisation and metabolic engineering. *Adv. Biochem. Eng. Biotechnol.* 148, 107–139. doi: 10.1007/10_2014_301
- Pelot, K. A., Chen, R., Hagelthorn, D. M., Young, C. A., Addison, J. B., Muchlinski, A., et al. (2018). Functional diversity of diterpene synthases in the biofuel crop switchgrass. *Plant Physiol.* 178, 54–71. doi: 10.1104/pp.18.00590
- Pelot, K. A., Hagelthorn, D. M., Addison, J. B., and Zerbe, P. (2017). Biosynthesis of the oxygenated diterpene nezuol in the medicinal plant *Isodon rubescens* is catalyzed by a pair of diterpene synthases. *PLoS One* 12, e0176507. doi: 10.1371/journal.pone.0176507
- Pelot, K. A., Hagelthorn, D. M., Hong, Y. J., Tantillo, D. J., and Zerbe, P. (2019). Diterpene synthase-catalyzed biosynthesis of distinct clerodane stereoisomers. *ChemBiochem* 20, 111–117. doi: 10.1002/cbic.201800580
- Pelot, K. A., Mitchell, R., Kwon, M., Hagelthorn, D. M., Wardman, J. F., Chiang, A., et al. (2016). Biosynthesis of the psychotropic plant diterpene salvinorin A: discovery and characterization of the *Salvia divinorum* clerodienyl diphosphate synthase. *Plant J.* 89, 885–897. doi: 10.1111/tpj.13427
- Peralta-Yahya, P. P., Zhang, F., Del Cardayre, S. B., and Keasling, J. D. (2012). Microbial engineering for the production of advanced biofuels. *Nature* 488, 320–328. doi: 10.1038/nature11478
- Peters, R. J. (2010). Two rings in them all: the labdane-related diterpenoids. *Nat. Prod. Rep.* 27, 1521–1530. doi: 10.1039/c0np00019a
- Peters, R. J., and Croteau, R. B. (2002). Abietadiene synthase catalysis: conserved residues involved in protonation-initiated cyclization of geranylgeranyl diphosphate to (+)-copalyl diphosphate. *Biochemistry* 41, 1836–1842. doi: 10.1021/bi011879d
- Peters, R. J., Flory, J. E., Jetter, R., Ravn, M. M., Lee, H. J., Coates, R. M., et al. (2000). Abietadiene synthase from grand fir (*Abies grandis*): characterization and mechanism of action of the “pseudomature” recombinant enzyme. *Biochemistry* 39, 15592–15602. doi: 10.1021/bi001997l
- Philippe, R. N., De Mey, M., Anderson, J., and Ajikumar, P. K. (2014). Biotechnological production of natural zero-calorie sweeteners. *Curr. Opin. Biotechnol.* 26, 155–161. doi: 10.1016/j.copbio.2014.01.004
- Pichersky, E., and Gershenzon, J. (2002). The formation and function of plant volatiles: perfumes for pollinator attraction and defense. *Curr. Opin. Plant Biol.* 5, 237–243. doi: 10.1016/S1369-5266(02)00251-0
- Pick, T. R., and Weber, A. P. (2014). Unknown components of the plastidial permeome. *Front. Plant Sci.* 5, 410. doi: 10.3389/fpls.2014.00410
- Potter, K., Criswell, J., Zi, J., Stubbs, A., and Peters, R. J. (2014). Novel product chemistry from mechanistic analysis of *ent*-copalyl diphosphate synthases from plant hormone biosynthesis. *Angew. Chem. Int. Ed. Engl.* 53, 7198–7202. doi: 10.1002/anie.201402911
- Potter, K. C., Jia, M., Hong, Y. J., Tantillo, D., and Peters, R. J. (2016a). Product rearrangement from altering a single residue in the rice *syn*-copalyl diphosphate synthase. *Org. Lett.* 18, 1060–1063. doi: 10.1021/acs.orglett.6b00181
- Potter, K. C., Zi, J., Hong, Y. J., Schulte, S., Malchow, B., Tantillo, D. J., et al. (2016b). Blocking deprotonation with retention of aromaticity in a plant *ent*-copalyl diphosphate synthase leads to product rearrangement. *Angew. Chem. Int. Ed. Engl.* 55, 634–638. doi: 10.1002/anie.201509060
- Prisic, S., Xu, J., Coates, R. M., and Peters, R. J. (2007). Probing the role of the DXDD motif in class II diterpene cyclases. *ChemBiochem* 8, 869–874. doi: 10.1002/cbic.200700045
- Qu, Y., Chakrabarty, R., Tran, H. T., Kwon, E. J., Kwon, M., Nguyen, T. D., et al. (2015). A lettuce (*Lactuca sativa*) homolog of human Nogo-B receptor interacts with *cis*-prenyltransferase and is necessary for natural rubber biosynthesis. *J. Biol. Chem.* 290, 1898–1914. doi: 10.1074/jbc.M114.616920
- Quin, M. B., Flynn, C. M., and Schmidt-Dannert, C. (2014). Traversing the fungal terpenome. *Nat. Prod. Rep.* 31, 1449–1473. doi: 10.1039/C4NP00075G
- Rivera, S. B., Swedlund, B. D., King, G. J., Bell, R. N., Hussey, C.E.Jr., Shattuck-Eidens, D. M., et al. (2001). Chrysanthemyl diphosphate synthase: isolation of the gene and characterization of the recombinant non-head-to-tail monoterpene synthase from *Chrysanthemum cinerariaefolium*. *Proc. Natl. Acad. Sci. U. S. A.* 98, 4373–4378. doi: 10.1073/pnas.071543598
- Ro, D. K., Arimura, G., Lau, S. Y., Piers, E., and Bohlmann, J. (2005). Loblolly pine abietadienol/abietadienal oxidase PtAO (CYP720B1) is a multifunctional, multisubstrate cytochrome P450 monooxygenase. *Proc. Natl. Acad. Sci. U. S. A.* 102, 8060–8065. doi: 10.1073/pnas.0500825102
- Ro, D. K., and Bohlmann, J. (2006). Diterpene resin acid biosynthesis in loblolly pine (*Pinus taeda*): functional characterization of abietadiene/levopimaradiene synthase (PtTPS-LAS) cDNA and subcellular targeting of PtTPS-LAS and abietadienol/abietadienal oxidase (PtAO, CYP720B1). *Phytochemistry* 67, 1572–1578. doi: 10.1016/j.phytochem.2006.01.011
- Roach, C. R., Hall, D. E., Zerbe, P., and Bohlmann, J. (2014). Plasticity and evolution of (+)-3-carene synthase and (-)-sabinene synthase functions of a sitka spruce monoterpene synthase gene family associated with weevil resistance. *J. Biol. Chem.* 289, 23859–23869. doi: 10.1074/jbc.M114.571703
- Rodriguez-Concepcion, M., and Boronat, A. (2015). Breaking new ground in the regulation of the early steps of plant isoprenoid biosynthesis. *Curr. Opin. Plant Biol.* 25, 17–22. doi: 10.1016/j.pbi.2015.04.001
- Rohmer, M., and Bissleret, P. (1994). Hopanoid and other polyterpenoid biosynthesis in Eubacteria—phylogenetic significance. *Isopentenoids Nat. Prod.* 562, 31–43. doi: 10.1021/bk-1994-0562.ch002
- Rontein, D., Onillon, S., Herbette, G., Lesot, A., Werck-Reichhart, D., Sallaud, C., et al. (2008). CYP725A4 from yew catalyzes complex structural rearrangement of taxa-4(5),11(12)-diene into the cyclic ether 5(12)-oxa-3(11)-cyclotaxane. *J. Biol. Chem.* 283, 6067–6075. doi: 10.1074/jbc.M708950200
- Rudolf, J. D., Dong, L. B., Cao, H. N., Hatzos-Skintges, C., Osipiuk, J., Endres, M., et al. (2016). Structure of the *ent*-copalyl diphosphate synthase PtmT2 from *Streptomyces platensis* CB00739, a bacterial type II diterpene synthase. *J. Am. Chem. Soc.* 138, 10905–10915. doi: 10.1021/jacs.6b04317
- Sallaud, C., Rontein, D., Onillon, S., Jabes, F., Duffe, P., Giacalone, C., et al. (2009). A novel pathway for sesquiterpene biosynthesis from Z,Z-farnesyl pyrophosphate in the wild tomato *Solanum habrochaites*. *Plant Cell* 21, 301–317. doi: 10.1105/tpc.107.057885
- Salmon, M., Laurendon, C., Vardakou, M., Cheema, J., Defernez, M., Green, S., et al. (2015). Emergence of terpene cyclization in *Artemisia annua*. *Nat. Commun.* 6, 6143. doi: 10.1038/ncomms7143
- Schalk, M., Pastore, L., Mirata, M. A., Khim, S., Schouwey, M., Deguerry, F., et al. (2012). Toward a biosynthetic route to sclareol and amber odorants. *J. Am. Chem. Soc.* 134, 18900–18903. doi: 10.1021/ja307404u
- Scheler, U., Brandt, W., Porzel, A., Rothe, K., Manzano, D., Božić, D., et al. (2016). Elucidation of the biosynthesis of carnosic acid and its reconstitution in yeast. *Nat. Commun.* 7, 12942. doi: 10.1038/ncomms12942

- Schepmann, H. G., Pang, J., and Matsuda, S. P. (2001). Cloning and characterization of *Ginkgo biloba* leupimaradiene synthase which catalyzes the first committed step in ginkgolide biosynthesis. *Arch. Biochem. Biophys.* 392, 263–269. doi: 10.1006/abbi.2001.2438
- Schillmiller, A. L., Schauvinhold, I., Larson, M., Xu, R., Charbonneau, A. L., Schmidt, A., et al. (2009). Monoterpenes in the glandular trichomes of tomato are synthesized from a neryl diphosphate precursor rather than geranyl diphosphate. *Proc. Natl. Acad. Sci. U. S. A.* 106, 10865–10870. doi: 10.1073/pnas.0904113106
- Schmelz, E. A., Huffaker, A., Sims, J. W., Christensen, S. A., Lu, X., Okada, K., et al. (2014). Biosynthesis, elicitation and roles of monocot terpenoid phytoalexins. *Plant J.* 79, 659–678. doi: 10.1111/tpj.12436
- Schmidt-Dannert, C. (2015). Biosynthesis of terpenoid natural products in fungi. *Adv. Biochem. Eng. Biotechnol.* 148, 19–61. doi: 10.1007/10_2014_283
- Schrepfer, P., Buettner, A., Goerner, C., Hertel, M., van Rijn, J., Wallrapp, F., et al. (2016). Identification of amino acid networks governing catalysis in the closed complex of class I terpene synthases. *Proc. Natl. Acad. Sci. U. S. A.* 113, E958–E967. doi: 10.1073/pnas.1519680113
- Schulte, S., Potter, K. C., Lemke, C., and Peters, R. J. (2018). Catalytic bases and stereocontrol in Lamiaceae class II diterpene cyclases. *Biochemistry* 57, 3473–3479. doi: 10.1021/acs.biochem.8b00193
- Shalev, T. J., Yuen, M. M. S., Gesell, A., Yuen, A., Russell, J. H., and Bohlmann, J. (2018). An annotated transcriptome of highly inbred *Thuja plicata* (Cupressaceae) and its utility for gene discovery of terpenoid biosynthesis and conifer defense. *Tree Genet. Genomes* 14, 35. doi: 10.1007/s11295-018-1248-y
- Shishova, E. Y., Yu, F., Miller, D. J., Faraldos, J. A., Zhao, Y., Coates, R. M., et al. (2008). X-ray crystallographic studies of substrate binding to aristolochene synthase suggest a metal ion binding sequence for catalysis. *J. Biol. Chem.* 283, 15431–15439. doi: 10.1074/jbc.M800659200
- Smanski, M. J., Peterson, R. M., Huang, S. X., and Shen, B. (2012). Bacterial diterpene synthases: new opportunities for mechanistic enzymology and engineered biosynthesis. *Curr. Opin. Chem. Biol.* 16, 132–141. doi: 10.1016/j.cbpa.2012.03.002
- Soler, E., Clastre, M., Bantignies, B., Marigo, G., and Ambid, C. (1993). Uptake of isopentenyl diphosphate by plastids isolated from *Vitis-Vinifera* L. cell-suspensions. *Planta* 191, 324–329. doi: 10.1007/BF00195689
- Srividya, N., Davis, E. M., Croteau, R. B., and Lange, B. M. (2015). Functional analysis of (4S)-limonene synthase mutants reveals determinants of catalytic outcome in a model monoterpene synthase. *Proc. Natl. Acad. Sci. U. S. A.* 112, 3332–3337. doi: 10.1073/pnas.1501203112
- Starks, C. M. (1997). Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase. *Science* 277, 1815–1820. doi: 10.1126/science.277.5333.1815
- Sugai, Y., Ueno, Y., Hayashi, K., Oogami, S., Toyomasu, T., Matsumoto, S., et al. (2011). Enzymatic (¹³C) labeling and multidimensional NMR analysis of mitraterpene synthesized by bifunctional diterpene cyclase in *Selaginella moellendorffii*. *J. Biol. Chem.* 286, 42840–42847. doi: 10.1074/jbc.M111.302703
- Swaminathan, S., Morrone, D., Wang, Q., Fulton, D. B., and Peters, R. J. (2009). CYP76M7 is an *ent*-cassadiene C11 α -hydroxylase defining a second multifunctional diterpenoid biosynthetic gene cluster in rice. *Plant Cell* 21, 3315–3325. doi: 10.1105/tpc.108.063677
- Tantillo, D. J. (2010). The carbocation continuum in terpene biosynthesis—where are the secondary cations? *Chem. Soc. Rev.* 39, 2847–2854. doi: 10.1039/b917107j
- Tantillo, D. J. (2011). Biosynthesis *via* carbocations: theoretical studies on terpene formation. *Nat. Prod. Rep.* 28, 1035–1053. doi: 10.1039/c1np00006c
- Tholl, D. (2006). Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. *Curr. Opin. Plant Biol.* 9, 297–304. doi: 10.1016/j.pbi.2006.03.014
- Tholl, D. (2015). Biosynthesis and biological functions of terpenoids in plants. *Adv. Biochem. Eng. Biotechnol.* 148, 63–106. doi: 10.1007/10_2014_295
- Tholl, D., Kish, C. M., Orlova, I., Sherman, D., Gershenzon, J., Pichersky, E., et al. (2004). Formation of monoterpenes in *Antirrhinum majus* and *Clarkia breweri* flowers involves heterodimeric geranyl diphosphate synthases. *Plant Cell* 16, 977–992. doi: 10.1105/tpc.020156
- Tian, B., Poulter, C. D., and Jacobson, M. P. (2016). Defining the product chemical space of monoterpene synthases. *PLoS Comput. Biol.* 12, e1005053. doi: 10.1371/journal.pcbi.1005053
- Tian, B. X., Wallrapp, F. H., Holiday, G. L., Chow, J. Y., Babbitt, P. C., Poulter, C. D., et al. (2014). Predicting the functions and specificity of triterpenoid synthases: a mechanism-based multi-intermediate docking approach. *PLoS Comput. Biol.* 10, e1003874. doi: 10.1371/journal.pcbi.1003874
- Toyomasu, T., Kawaide, H., Ishizaki, A., Shinoda, S., Otsuka, M., Mitsunashi, W., et al. (2000). Cloning of a full-length cDNA encoding *ent*-kaurene synthase from *Gibberella fujikuroi*: functional analysis of a bifunctional diterpene cyclase. *Biosci. Biotechnol. Biochem.* 64, 660–664. doi: 10.1271/bbb.64.660
- Toyomasu, T., Nakaminami, K., Toshima, H., Mie, T., Watanabe, K., Ito, H., et al. (2004). Cloning of a gene cluster responsible for the biosynthesis of diterpene aphidicolin, a specific inhibitor of DNA polymerase α . *Biosci. Biotechnol. Biochem.* 68, 146–152. doi: 10.1271/bbb.68.146
- Toyomasu, T., Niida, R., Kenmoku, H., Kanno, Y., Miura, S., Nakano, C., et al. (2008). Identification of diterpene biosynthetic gene clusters and functional analysis of labdane-related diterpene cyclases in *Phomopsis amygdali*. *Biosci. Biotechnol. Biochem.* 72, 1038–1047. doi: 10.1271/bbb.70790
- Toyomasu, T., Tsukahara, M., Kaneko, A., Niida, R., Mitsunashi, W., Dai, T., et al. (2007). Fusicoccins are biosynthesized by an unusual chimera diterpene synthase in fungi. *Proc. Natl. Acad. Sci. U. S. A.* 104, 3084–3088. doi: 10.1073/pnas.0608426104
- Trapp, S. C., and Croteau, R. B. (2001). Genomic organization of plant terpene synthases and molecular evolutionary implications. *Genetics* 158, 811–832.
- Tripathy, B. C., and Pattanayak, G. K. (2012). Chlorophyll biosynthesis in higher plants. *Photosyn. Plastid Biol. Energy Convers. Carbon Assimilation* 34, 63–94. doi: 10.1007/978-94-007-1579-0_3
- Tudzynski, B. (2005). Gibberellin biosynthesis in fungi: genes, enzymes, evolution, and impact on biotechnology. *Appl. Microbiol. Biotechnol.* 66, 597–611. doi: 10.1007/s00253-004-1805-1
- Tudzynski, B., and Holter, K. (1998). Gibberellin biosynthetic pathway in *Gibberella fujikuroi*: evidence for a gene cluster. *Fungal Genet. Biol.* 25, 157–170. doi: 10.1006/fgbi.1998.1095
- Van De Vossenberg, J. L., Driessen, A. J., and Konings, W. N. (1998). The essence of being extremophilic: the role of the unique archaeal membrane lipids. *Extremophiles* 2, 163–170. doi: 10.1007/s007920050056
- Vannice, J. C., Scaff, D. A., Keightley, A., Addo, J. K., Wyckoff, G. J., and Mizioro, H. M. (2014). Identification in *Haloferax volcanii* of phosphomevalonate decarboxylase and isopentenyl phosphate kinase as catalysts of the terminal enzyme reactions in an archaeal alternate mevalonate pathway. *J. Bacteriol.* 196, 1055–1063. doi: 10.1128/JB.01230-13
- Vaughan, M. M., Christensen, S., Schmelz, E. A., Huffaker, A., McAuslane, H. J., Alborn, H. T., et al. (2015). Accumulation of terpenoid phytoalexins in maize roots is associated with drought tolerance. *Plant Cell Environ.* 38, 2195–2207. doi: 10.1111/pce.12482
- Vaughan, M. M., Wang, Q., Webster, F. X., Kiemle, D., Hong, Y. J., Tantillo, D. J., et al. (2013). Formation of the unusual semivolatile diterpene rhizathalene by the Arabidopsis class I terpene synthase TPS08 in the root stele is involved in defense against belowground herbivory. *Plant Cell* 25, 1108–1125. doi: 10.1105/tpc.112.100057
- Vranova, E., Coman, D., and Gruissem, W. (2013). Network analysis of the MVA and MEP pathways for isoprenoid synthesis. *Annu. Rev. Plant Biol.* 64, 665–700. doi: 10.1146/annurev-arplant-050312-120116
- Wang, Q., Hillwig, M. L., Okada, K., Yamazaki, K., Wu, Y., Swaminathan, S., et al. (2012a). Characterization of CYP76M5-8 indicates metabolic plasticity within a plant biosynthetic gene cluster. *J. Biol. Chem.* 287, 6159–6168. doi: 10.1074/jbc.M111.305599
- Wang, Q., Hillwig, M. L., Wu, Y., and Peters, R. J. (2012b). CYP701A8: a rice *ent*-kaurene oxidase paralog diverted to more specialized diterpenoid metabolism. *Plant Physiol.* 158, 1418–1425. doi: 10.1104/pp.111.187518
- Wei, G., Jia, Q., Chen, X., Köllner, T. G., Bhattacharya, D., Wong, G. K., et al. (2018). Terpene biosynthesis in red algae is catalyzed by microbial type but not typical plant terpene synthases. *Plant Physiol.* 179, 382–390. doi: 10.1104/pp.18.01413
- Whittington, D. A., Wise, M. L., Urbansky, M., Coates, R. M., Croteau, R. B., and Christianson, D. W. (2002). Bornyl diphosphate synthase: structure and strategy for carbocation manipulation by a terpenoid cyclase. *Proc. Natl. Acad. Sci. U. S. A.* 99, 15375–15380. doi: 10.1073/pnas.232591099
- Wilderman, P. R., and Peters, R. J. (2007). A single residue switch converts abietadiene synthase into a pimaradiene specific cyclase. *J. Am. Chem. Soc.* 129, 15736–15737. doi: 10.1021/ja074977g

- Wilderman, P. R., Xu, M., Jin, Y., Coates, R. M., and Peters, R. J. (2004). Identification of *syn*-pimara-7,15-diene synthase reveals functional clustering of terpene synthases involved in rice phytoalexin/allelochemical biosynthesis. *Plant Physiol.* 135, 2098–2105. doi: 10.1104/pp.104.045971
- Wille, A., Zimmermann, P., Vranova, E., Furholz, A., Laule, O., Bleuler, S., et al. (2004). Sparse graphical Gaussian modeling of the isoprenoid gene network in *Arabidopsis thaliana*. *Genome Biol.* 5, R92. doi: 10.1186/gb-2004-5-11-r92
- Williams, D. C., Wildung, M. R., Jin, A. Q., Dalal, D., Oliver, J. S., Coates, R. M., et al. (2000). Heterologous expression and characterization of a “Pseudomature” form of taxadiene synthase involved in paclitaxel (Taxol) biosynthesis and evaluation of a potential intermediate and inhibitors of the multistep diterpene cyclization reaction. *Arch. Biochem. Biophys.* 379, 137–146. doi: 10.1006/abbi.2000.1865
- Wu, Y., Hillwig, M. L., Wang, Q., and Peters, R. J. (2011). Parsing a multifunctional biosynthetic gene cluster from rice: biochemical characterization of CYP71Z6 & 7. *FEBS Lett.* 585, 3446–3451. doi: 10.1016/j.febslet.2011.09.038
- Wu, Y., Zhou, K., Toyomasu, T., Sugawara, C., Oku, M., Abe, S., et al. (2012). Functional characterization of wheat copalyl diphosphate synthases sheds light on the early evolution of labdane-related diterpenoid metabolism in the cereals. *Phytochemistry* 84, 40–46. doi: 10.1016/j.phytochem.2012.08.022
- Wurtzel, E. T., and Kutchan, T. M. (2016). Plant metabolism, the diverse chemistry set of the future. *Science* 353, 1232–1236. doi: 10.1126/science.aad2062
- Xiong, W., Fu, J., Köllner, T. G., Chen, X., Jia, Q., Guo, H., et al. (2018). Biochemical characterization of microbial type terpene synthases in two closely related species of hornworts, *Anthoceros punctatus* and *Anthoceros agrestis*. *Phytochemistry* 149, 116–122. doi: 10.1016/j.phytochem.2018.02.011
- Xu, M., Hillwig, M. L., Priscic, S., Coates, R. M., and Peters, R. J. (2004). Functional identification of rice *syn*-copalyl diphosphate synthase and its role in initiating biosynthesis of diterpenoid phytoalexin/allelopathic natural products. *Plant J.* 39, 309–318. doi: 10.1111/j.1365-313X.2004.02137.x
- Xu, M., Wilderman, P. R., Morrone, D., Xu, J., Roy, A., Margis-Pinheiro, M., et al. (2007a). Functional characterization of the rice kaurene synthase-like gene family. *Phytochemistry* 68, 312–326. doi: 10.1016/j.phytochem.2006.10.016
- Xu, M., Wilderman, P. R., and Peters, R. J. (2007b). Following evolution's lead to a single residue switch for diterpene synthase product outcome. *Proc. Natl. Acad. Sci. U. S. A.* 104, 7397–7401. doi: 10.1073/pnas.0611454104
- Yamada, Y., Cane, D. E., and Ikeda, H. (2012). Diversity and analysis of bacterial terpene synthases. *Methods Enzymol.* 515, 123–162. doi: 10.1016/B978-0-12-394290-6.00007-0
- Zager, J. J., Lange, I., Srividya, N., Smith, A., and Lange, B. M. (2019). Gene networks underlying cannabinoid and terpenoid accumulation in cannabis. *Plant Physiol.* 180, 1877–1897. doi: 10.1104/pp.18.01506
- Zerbe, P., and Bohlmann, J. (2015). Plant diterpene synthases: exploring modularity and metabolic diversity for bioengineering. *Trends Biotechnol.* 33, 419–428. doi: 10.1016/j.tibtech.2015.04.006
- Zerbe, P., Chiang, A., and Bohlmann, J. (2012a). Mutational analysis of white spruce (*Picea glauca*) *ent*-kaurene synthase (PgKS) reveals common and distinct mechanisms of conifer diterpene synthases of general and specialized metabolism. *Phytochemistry* 74, 30–39. doi: 10.1016/j.phytochem.2011.11.004
- Zerbe, P., Chiang, A., Dullat, H., O'Neil-Johnson, M., Starks, C., Hamberger, B., et al. (2014). Diterpene synthases of the biosynthetic system of medicinally active diterpenoids in *Marrubium vulgare*. *Plant J.* 79, 914–927. doi: 10.1111/tpj.12589
- Zerbe, P., Chiang, A., Yuen, M., Hamberger, B., Hamberger, B., Draper, J. A., et al. (2012b). Bifunctional *cis*-abienol synthase from *Abies balsamea* discovered by transcriptome sequencing and its implications for diterpenoid fragrance production. *J. Biol. Chem.* 287, 12121–12131. doi: 10.1074/jbc.M111.317669
- Zerbe, P., Hamberger, B., Yuen, M. M., Chiang, A., Sandhu, H. K., Madilao, L. L., et al. (2013). Gene discovery of modular diterpene metabolism in nonmodel systems. *Plant Physiol.* 162, 1073–1091. doi: 10.1104/pp.113.218347
- Zerbe, P., Rodriguez, S. M., Mafu, S., Chiang, A., Sandhu, H. K., O'Neil-Johnson, M., et al. (2015). Exploring diterpene metabolism in non-model species: transcriptome-enabled discovery and functional characterization of labda-7,13E-dienyl diphosphate synthase from *Grindelia robusta*. *Plant J.* 83, 783–793. doi: 10.1111/tpj.12925
- Zhou, K., Gao, Y., Hoy, J. A., Mann, F. M., Honzatko, R. B., and Peters, R. J. (2012a). Insights into diterpene cyclization from structure of bifunctional abietadiene synthase from *Abies grandis*. *J. Biol. Chem.* 287, 6840–6850. doi: 10.1074/jbc.M111.337592
- Zhou, K., Xu, M., Tiernan, M., Xie, Q., Toyomasu, T., Sugawara, C., et al. (2012b). Functional characterization of wheat *ent*-kaurene(-like) synthases indicates continuing evolution of labdane-related diterpenoid metabolism in the cereals. *Phytochemistry* 84, 47–55. doi: 10.1016/j.phytochem.2012.08.021
- Zi, J., Mafu, S., and Peters, R. J. (2014). To gibberellins and beyond! Surveying the evolution of (di)terpenoid metabolism. *Annu. Rev. Plant Biol.* 65, 259–286. doi: 10.1146/annurev-arplant-050213-035705

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Karunanithi and Zerbe. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Taxonomically Informed Scoring Enhances Confidence in Natural Products Annotation

Adriano Rutz^{1†}, Miwa Dounoue-Kubo^{1,2†}, Simon Ollivier¹, Jonathan Bisson^{3,4}, Mohsen Bagheri^{1,5}, Tongchai Saesong⁶, Samad Nejad Ebrahimi⁵, Kornkanok Ingkaninan⁶, Jean-Luc Wolfender^{1*} and Pierre-Marie Allard^{1*}

¹ Institute of Pharmaceutical Sciences of Western Switzerland (ISPSO), University of Geneva, Centre Médical Universitaire (CMU), Geneva, Switzerland, ² Faculty of Pharmaceutical Sciences, Tokushima Bunri University, Tokushima, Japan, ³ Center for Natural Product Technologies, Program for Collaborative Research in the Pharmaceutical Sciences (PCRPS), University of Illinois at Chicago, Chicago, IL, United States, ⁴ Department of Pharmaceutical Sciences, College of Pharmacy, University of Illinois at Chicago, Chicago, IL, United States, ⁵ Department of Phytochemistry, Medicinal Plants and Drugs Research Institute, Shahid Beheshti University, G.C., Evin, Tehran, Iran, ⁶ Department of Pharmaceutical Chemistry and Pharmacognosy, Faculty of Pharmaceutical Sciences and Center of Excellence for Innovation in Chemistry, Naresuan University, Phitsanulok, Thailand

OPEN ACCESS

Edited by:

Kazuki Saito,
RIKEN Center for Sustainable
Resource Science (CSRS),
Japan

Reviewed by:

Hiroshi Tsugawa,
RIKEN, Japan
Tobias Kind,
University of California, Davis,
United States

*Correspondence:

Jean-Luc Wolfender
jean-luc.wolfender@unige.ch
Pierre-Marie Allard
pierre-marie.allard@unige.ch

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Plant Metabolism and
Chemodiversity,
a section of the journal
Frontiers in Plant Science

Received: 14 July 2019

Accepted: 24 September 2019

Published: 25 October 2019

Citation:

Rutz A, Dounoue-Kubo M, Ollivier S,
Bisson J, Bagheri M, Saesong T,
Ebrahimi SN, Ingkaninan K,
Wolfender J-L and Allard P-M (2019)
Taxonomically Informed Scoring
Enhances Confidence in Natural
Products Annotation.
Front. Plant Sci. 10:1329.
doi: 10.3389/fpls.2019.01329

Mass spectrometry (MS) offers unrivalled sensitivity for the metabolite profiling of complex biological matrices encountered in natural products (NP) research. The massive and complex sets of spectral data generated by such platforms require computational approaches for their interpretation. Within such approaches, computational metabolite annotation automatically links spectral data to candidate structures *via* a score, which is usually established between the acquired data and experimental or theoretical spectral databases (DB). This process leads to various candidate structures for each MS features. However, at this stage, obtaining high annotation confidence level remains a challenge notably due to the extensive chemodiversity of specialized metabolomes. The design of a metascore is a way to capture complementary experimental attributes and improve the annotation process. Here, we show that integrating the taxonomic position of the biological source of the analyzed samples and candidate structures enhances confidence in metabolite annotation. A script is proposed to automatically input such information at various granularity levels (species, genus, and family) and complement the score obtained between experimental spectral data and output of available computational metabolite annotation tools (ISDB-DNP, MS-Finder, Sirius). In all cases, the consideration of the taxonomic distance allowed an efficient re-ranking of the candidate structures leading to a systematic enhancement of the recall and precision rates of the tools (1.5- to 7-fold increase in the F1 score). Our results clearly demonstrate the importance of considering taxonomic information in the process of specialized metabolites annotation. This requires to access structural data systematically documented with biological origin, both for new and previously reported NPs. In this respect, the establishment of an open structural DB of specialized metabolites and their associated metadata, particularly biological sources, is timely and critical for the NP research community.

Keywords: metabolite annotation, chemotaxonomy, scoring system, natural products, computational metabolomics, taxonomic distance, specialized metabolome

INTRODUCTION

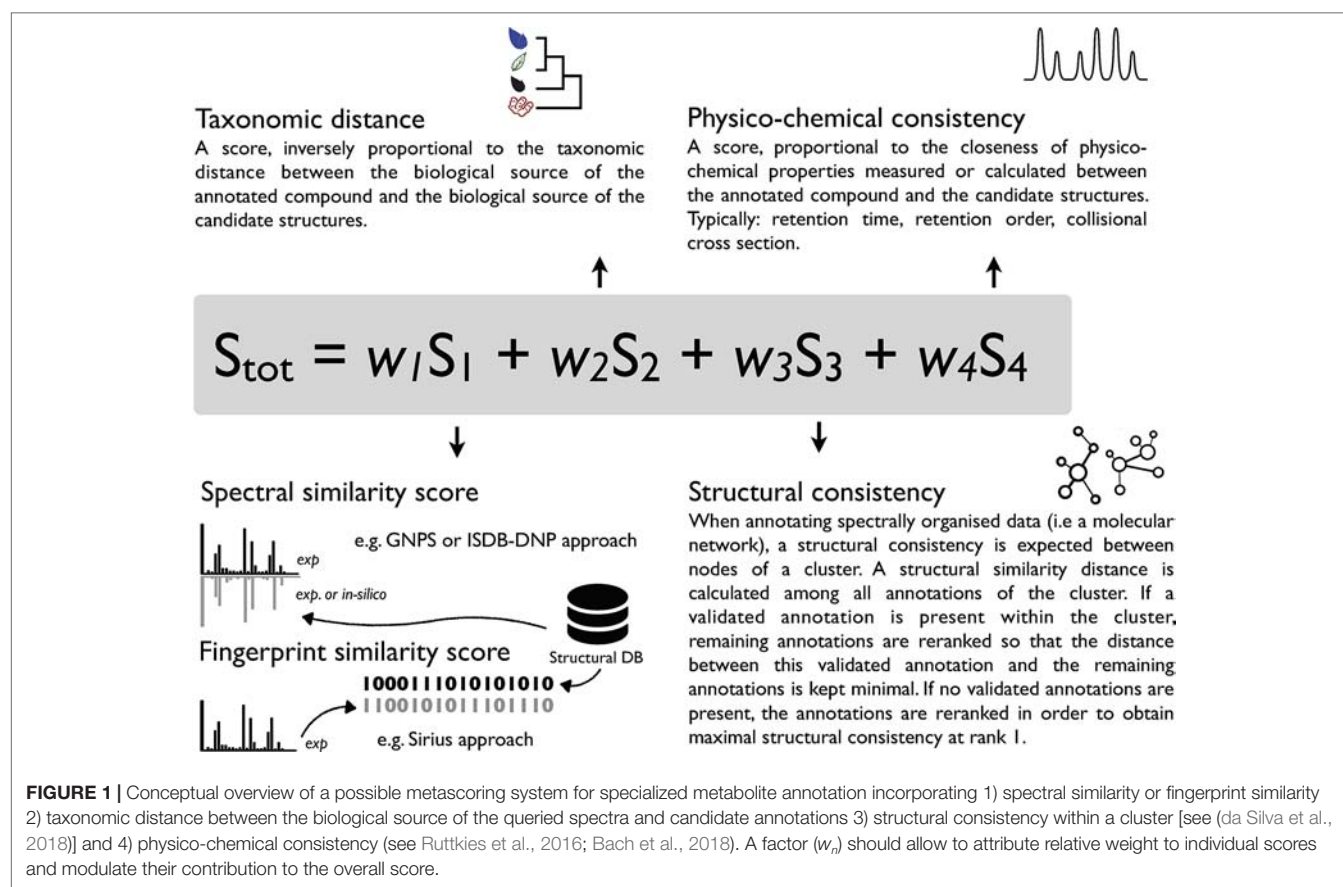
Specialized metabolites define the chemical signature of a living organism. Plants, sponges and corals, but also microorganisms (bacteria and fungi), are known to biosynthesize a wealth of such chemicals, which can play a role as defense or communication agents (Brunetti et al., 2018). Throughout history, humans have been relying on plant derived products for a variety of purposes: housing, feeding, clothing and, especially, medication. In fact, our therapeutic arsenal is deeply dependent on the chemistry of natural products (NPs) whether they are used in mixtures, purified forms or for hemi-synthetic drug development. After a period of disregard by the pharmaceutical industry, NPs are now the object of renewed interest, partly because of the promises of the latest technological developments (Shen, 2015). Developments in metabolite profiling by mass spectrometry (MS) grant access to large volumes of high-quality spectral data from minimal amount of samples and appropriate data analysis workflows allow to efficiently mine such data (Wolfender et al., 2019). Initiatives such as the Global Natural Products Social (GNPS) molecular networking (MN) project offer both a living MS repository and the possibility to establish MN organizing MS data (Wang et al., 2016). However, despite such advancements, *metabolite identification* remains a major challenge for both NP research and metabolomics (Kind et al., 2018). Metabolite identification of a novel compound requires physical isolation of the analyte followed by complete NMR acquisition and three-dimensional structural establishment *via* X-ray diffraction or chiroptical techniques. For previously described compounds, metabolite identification implies complete matching of physicochemical properties between the analyte and a standard compound (including chiroptical properties). Metabolite identification is thus a tedious and labor-intensive process, which should ideally be reserved to novel metabolites description. Any less complete process should be defined as *metabolite annotation*. By definition, metabolite annotation can be applied at a higher throughput and offers an effective proxy for the chemical characterization of complex matrices. This process includes dereplication (the annotation of previously described molecules prior to any physical isolation process) and allows focusing isolation and metabolite identification efforts on potentially novel compounds only (Gaudêncio and Pereira, 2015).

Given its sensitivity, selectivity and structural determination potential, MS is a tool of choice for metabolite annotation in complex mixtures. Various computational MS solutions have been developed to link experimental spectra to chemical structures. They can be classified into experimental rule-based strategies [MassHunter, Agilent Technologies], combinatorial fragmentation strategies [MetFrag, (Ruttkies et al., 2016)], machine learning based approaches using stochastic Markov modelling [CFM-ID, (Allen et al., 2014; Djoumbou-Feunang et al., 2019)] or predicting fragmentation trees [Sirius (Böcker et al., 2009; Dührkop et al., 2019)]. Computationally demanding *ab initio* calculations, modeling the gas-phase fragmentation process, have also been proposed (Bauer and Grimme, 2016). The output of such tools is, in general, a list of candidate molecules ranked according to a score. Such score can be based on a single measure (e.g. spectral similarity in CFM-annotate)

(Allen et al., 2014) or integrate combined parameters (MS-Finder, Sirius) (Tsugawa et al., 2016; Dührkop et al., 2019; Tsugawa et al., 2019). The rationale behind comprehensive scoring systems is that orthogonal information (not directly related to spectral comparison) should further strengthen the metabolite annotation process. This has been illustrated in the past by using the number of literature references related to a candidate structure and basic retention time scoring based on logP in MetFrag 2.2 (Ruttkies et al., 2016). Recently, the integration of retention order prediction to an MS/MS prediction tool provided increased performance in metabolite annotation (Bach et al., 2018). Another example is the Network Annotation Propagation (NAP) approach, which takes advantage of the topology of a MN to proceed to a re-ranking of annotated candidates within a cluster where structural consistency is expected (da Silva et al., 2018). In our view, increased confidence in specialized metabolite annotation can be achieved by the establishment of a metascore system capturing the similarity of diverse attributes shared by the queried analytes and candidate structures (Allard et al., 2017). Such metascore could for example consider 1) spectral similarity 2) taxonomic distance between the producer of the candidate compound and the annotated biological matrix 3) structural consistency within a cluster and 4) physico-chemical consistency. A conceptual overview of such metascore is illustrated in **Figure 1**. To the best of our knowledge, the automated inclusion of the taxonomic dimension within a scoring system has not been considered in current metabolite annotation strategies.

The central hypothesis of this work is directly inferred from the characteristics of the specialized metabolome. Unlike primary metabolites, which are mostly ubiquitous compounds central to organism functioning, specialized metabolites are, by definition, strongly linked to the taxonomic position of the producing organism. It thus appears desirable to consider taxonomic information when describing the chemistry of an organism. A *taxonomic filtering* process could be used to limit a database (DB) to compounds previously isolated from organisms situated within a given taxonomic distance from the biological source of the analyte to annotate. However, results of chemotaxonomic studies also highlight the presence of broadly distributed metabolites. For example, liriodenine (MUMCCPUVOAUBAN-UHFFFAOYSA-N) is a widely distributed alkaloid produced by more than 50 distinct biological sources, it is found in over 30 genera belonging to 13 botanical families. Convergent biosynthetic pathways offer intriguing example of unrelated species, shaped by evolution, that end up producing similar classes of compounds (Pichersky and Lewinsohn, 2011). To proceed to the annotation of such compounds, a *taxonomically informed scoring* allowing, both, to consider spectral similarity and taxonomic information while conserving the independence of the individual resulting scores appears as a better solution than a basic filtering process.

In the frame of this study we propose such taxonomically informed scoring system and benchmark the impact of taxonomic distance consideration on a set of 2,107 identified molecules using three different computational mass spectrometry metabolite annotation tools (ISDB-DNP, MS-Finder and Sirius).



RESULTS

Conception of the Taxonomically Informed Scoring System

The constituents of specialized metabolomes, as expression products of the genome, should reflect the taxonomic position of the producing organisms. The initial hypothesis of this work is that *the attribution of a score reflecting the taxonomic distance between the biological source of the queried analyte and the one of the candidate structures, is a valuable input for a metabolite annotation process.*

Taxonomically informed scoring is proposed to complement the initial score (S_1 in **Figure 1**) attributed to candidate structures by existing metabolite annotation tool. To this end, the initial score is first normalized. Then, scores, inversely proportional to the taxa level difference (family < genus < species) are attributed when an exact match is observed between biological source denominations at the different taxa levels. The score corresponding to the shortest taxonomic distance is then added to the initial score. Candidates are further re-ranked according to the newly complemented score. In this study, no phylogenetic distances within taxa (e.g. family, genus or species) were considered due to high computational requirements, but the development of such an approach would be of interest. The general outline of the *taxonomically informed scoring* system is presented in **Figure 2**.

In order to apply the *taxonomically informed scoring* in a generic manner, the initial scores given by the metabolite annotation tools

were rescaled to obtain values ranging from 0 (worst candidate) to 1 (best candidate). The scores, given according to the taxonomic distance between the biological source of the queried spectra and the one of the candidate compounds, were integrated in the final score by a sum. This choice allows to keep independence between individual components of the metascore (see **Figure 1**). Since the boundaries of the candidates' normalized score in a given dataset are defined (0 to 1), the minimal score to be applied to the worst candidate for it to be ranked at the first position after *taxonomically informed scoring* is 1. Following our initial hypothesis, a score of 1 was thus given if a match between biological sources was found at the family taxa level. In the case where the initial maximal score (1) would be given to a candidate and added to a score corresponding to a match at the family level (1), a score of at least 2 should be given for a candidate having the worst score to be ranked above. A score of 2 was thus given if a match between biological sources was found at the genus taxa level. Following the same logic, a score of 3 was given for matches between biological sources at the species level.

Benchmarking the Influence of Taxonomically Informed Scoring in Metabolite Annotation

Establishment of a Benchmarking Dataset

In order to establish the importance of considering taxonomic information in metabolite annotation, an experimental reference dataset constituted by molecular structures, their MS/MS spectra

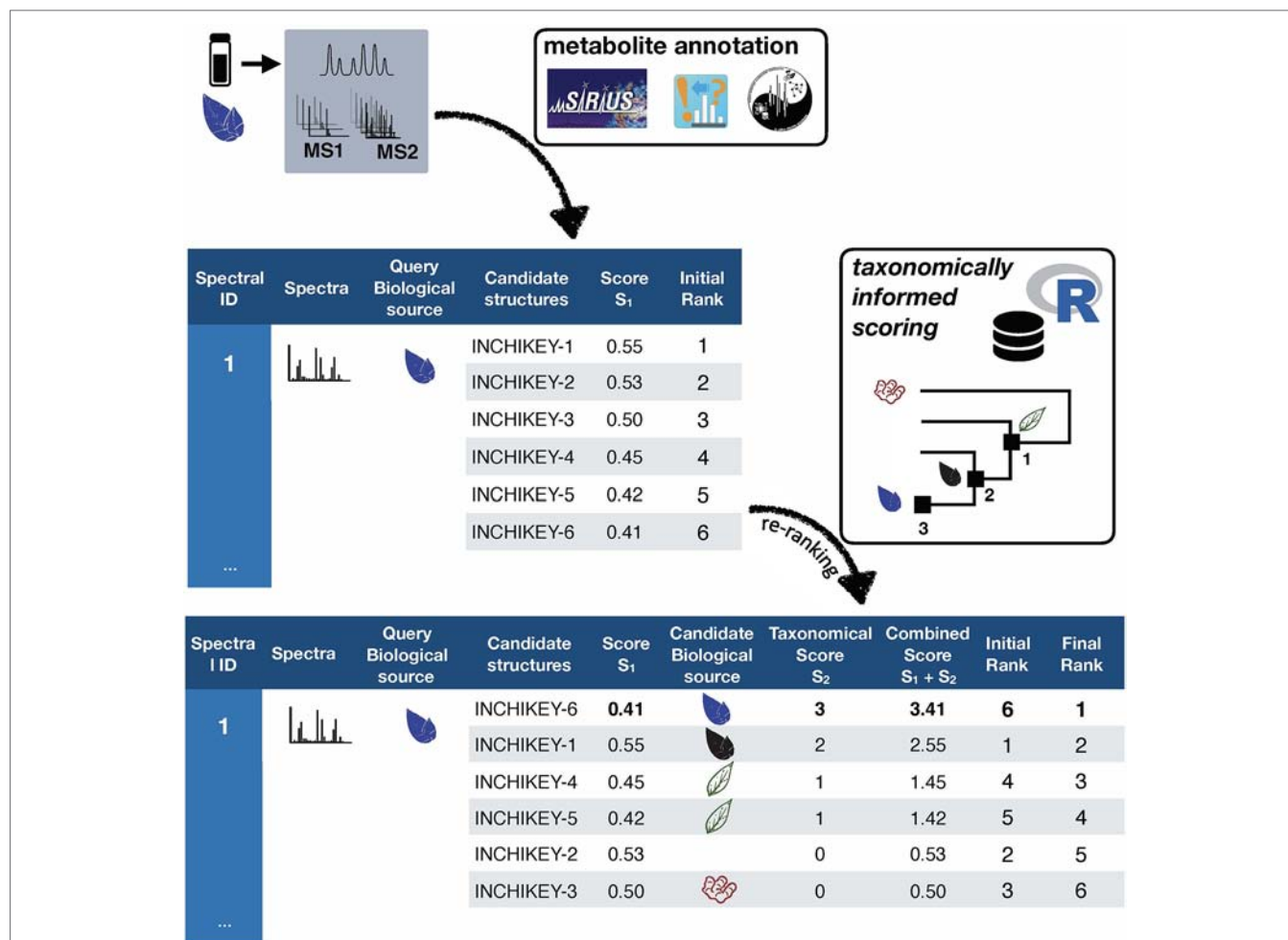


FIGURE 2 | General outline of the *taxonomically informed scoring* system. Candidates' structures are complemented with their biological sources at the family, genus and species level, when available. A score, inversely proportional to the taxonomic distance between the biological source of the standard compound and the one of the candidate compounds is given when the biological source of the candidate structures matches the biological source of the standard at the family, genus and species level, respectively. The maximum score for each candidate is then added to its spectral score to yield a complemented spectral score. Finally, candidates are re-ranked according to the complemented spectral score.

acquired under various experimental conditions and their biological sources, in the form of a fully resolved taxonomic hierarchy, was needed. This dataset, denominated hereafter *benchmarking dataset*, was built by combining a curated structural/biological sources dataset (obtained from the Dictionary of Natural Products (DNP)) and a curated structural/spectral dataset (obtained from GNPS libraries). Steps followed for the establishment of the benchmarking dataset are detailed below and summarized in **Supplementary Figure S3**.

Structural and Biological Sources Dataset

The prerequisite to apply a *taxonomically informed scoring* in a metabolite annotation process is to dispose of the biological source information of 1) the queried MS/MS spectra and 2) the candidate structures. To the best of our knowledge, there is currently no freely available database (DB) compiling NP structures and their biological sources down to the species level. This study uses the DNP which is commercially available and allows export of structures and biological sources as associated metadata. A curation

process using the Global Names index, kept biological sources resolved against the Catalogue of Life and resulted in 219,800 entries with accepted scientific names and a full, homogeneous, taxonomy up to the kingdom level. For example, the entry initially corresponding to pulsaquinone, "Constit. of the roots of *Pulsatilla koreana*", is converted to Plantae | Tracheophyta | Magnoliopsida | Ranunculales | Ranunculaceae | *Pulsatilla* | *Pulsatilla cernua* in the curated DB. See Material and Methods and **Supplementary Figure S3** for details concerning the curation process.

Structural and Spectral Dataset

The GNPS libraries agglomerate a wide and publicly available ensemble of MS/MS spectra coming from various analytical platforms and thus having different levels of quality (Wang et al., 2016). These spectral libraries were used as representative source of diverse experimental MS/MS spectra to evaluate the annotation improvement that could be obtained by applying *taxonomically informed scoring*. All GNPS libraries and publicly

accessible third-party libraries were retrieved online (<https://gnps.ucsd.edu/ProteoSAFe/libraries.jsp>) and concatenated as a single spectral file containing 66,646 individual entries. The pretreatment described in Material and Methods, yielded a dataset of 40,138 structures (8,558 unique structures) with their experimental associated MS/MS acquired on different platforms. See **Supplementary Figure S3**.

Structural, Spectral and Biological Sources Dataset (Benchmarking Dataset)

To apply the *taxonomically informed scoring*, it is required that denominations of both 1) the queried spectra and 2) the candidate structures biological sources are resolved using a common taxonomy backbone (i.e. using the accepted denomination). It was thus necessary to build an experimental spectral dataset for which each entry had a unique structure and a properly documented biological source, which constituted the benchmarking set. The structural and spectral dataset was matched against the structural and biological sources dataset, following the procedure detailed in Material and Methods. The full processing resulted in a dataset of 2,107 individual entries (characterized NPs with no stereoisomers distinction and a unique biological source associated), which was used for the rest of this study. See **Supplementary Figure S3**.

Analysis of the benchmarking dataset showed a chemodiversity comparable to the one of DNP (see panels A and C in **Figure 3**). Regarding the distribution of the biological sources in the benchmarking dataset, available data mostly matched plant specialized metabolites (see panel B in **Figure 3**). Additionally, repartition of mass analyzer types indicated the heterogeneous spectral quality of MS/MS spectra of the

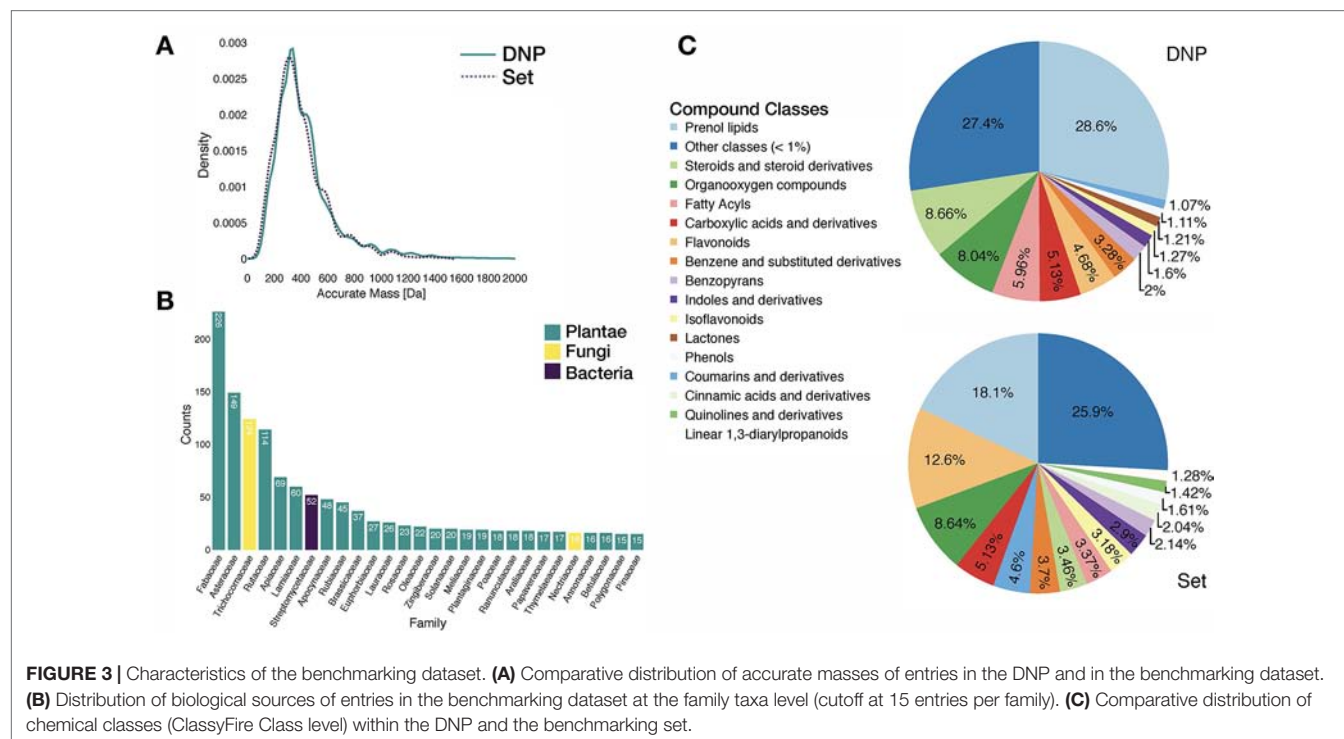
benchmarking dataset and was representative of commonly used analytical platforms. See repartition in **Supplementary Figure S7**.

Evaluation of the Improvement of Metabolite Annotation on the Benchmarking Set

In order to assess the importance of considering taxonomic information in the annotation process, the outputs of three different computational MS-based metabolite annotation solutions were considered (ISDB-DNP, MS-Finder and Sirius). The 2,107 spectra of the benchmarking dataset were queried using these tools. The precision and accuracy of structural determination with and without the use of *taxonomically informed scoring* were systematically compared according to parameters detailed in the Material and Methods section.

Metabolite Annotation Tools Used Isdb-Dnp

The first tool, denominated hereafter ISDB-DNP (*In Silico* DataBase—Dictionary of Natural Products) is a metabolite annotation strategy that we previously developed (Allard et al., 2016). This approach is focused on specialized metabolites annotation and is constituted by a pre-fragmented theoretical spectral DB version of the DNP. The *in silico* fragmentation was performed by CFM-ID (Allen et al., 2015). CFM-ID is, to our knowledge, the only computational solution currently available able to generate a theoretical spectrum with prediction of fragment intensity. The matching phase between experimental spectra and the theoretical DB is based on a spectral similarity measure (cosine score) performed using Tremolo (Wang and Bandeira, 2013). The scores are reported from 0 (worst candidate) to 1 (best candidate).



Ms-Finder. The second tool is MS-Finder. This *in silico* fragmentation approach considers multiple parameters such as bond dissociation energies, mass accuracies, fragment linkages and various hydrogen rearrangement rules at the candidate ranking phase (Tsugawa et al., 2016). The resulting scoring system range from 1 (worst candidate) to 10 (best candidate).

Sirius. The third tool to be used is Sirius 4.0. It is considered as a state-of-the-art metabolite annotation solution, which combines molecular formula calculation and the prediction of a molecular fingerprint of a query from its fragmentation tree and spectrum (Dührkop et al., 2019). Sirius uses a DB of 73,444,774 unique structures for its annotations. The resulting score is a probabilistic measure ranging between negative infinity (worst candidate) and 0 (best candidate).

Computation of the Taxonomically Informed Score

R scripts were written to perform 1) cleaning and standardization of the outputs, 2) *taxonomically informed scoring* and re-ranking. First, the outputs were standardized to a table containing on each row: the unique spectral identifier (CCCMSLIB N°) of the queried spectra, the short InChIKey of the candidate structures, the score of the candidates (within the scoring system of the used metabolite annotation tool), the biological source of the standard compound and the biological source of the candidate structures. As described above, a score, inversely proportional to the taxonomic distance between the biological source of the annotated compound and the biological source of the candidate structure, was given when both matched at the family (score of 1), genus (score of 2) or species level(s) (score of 3). A sum of this score (1 to 3) and the original score (0 to 1) yielded the *taxonomically informed score*. This *taxonomically informed score* was then used to re-rank the candidates from highest to lowest score.

Results Before Taxonomically Informed Scoring

Using each tools' initial scoring system, on the 2,107 experimental MS/MS spectra constituting the benchmarking set, the ISDB-DNP returned 214 (10.2%) correct annotations at rank 1, Sirius 975 (46.3%) and MS-Finder 180 (8.5%). The total number of unique correct annotations ranked first covered by ISDB-DNP, Sirius and MS-Finder prior to *taxonomically informed scoring* reached 1110 or 52.7% of the benchmarked dataset. Out of these, 29 (less than 1.4%), were common to all three tools, indicating the interest of considering various annotation tools when proceeding to metabolite annotation. Venn diagram in Figure 4 illustrates the complementarity of returned annotations. Within all candidates (all ranks), the ISDB-DNP returned 1,750 correct annotations, Sirius 1,589 and MS-Finder 574. The ROC curves outline the number of correct hits outside first rank and indicate remaining improvement potential. See Supplementary Figure S4.

Results After Taxonomically Informed Scoring

After *taxonomically informed scoring* and reranking, the number of correct annotations at rank 1 increased to 1,510, 1,508 and 546, respectively for ISDB-DNP, Sirius and MS-Finder. The

total number of correct annotations covered by all ISDB-DNP, Sirius and MS-Finder after *taxonomically informed scoring* reached 1786 or 84.8% of the benchmarked dataset. Interestingly, more than 10-fold increase after *taxonomically informed scoring* was also observed for the correctly annotated metabolite commonly returned by the three tools 376 (17%). It has to be noted that no stereoisomer distinction could be performed since all correct matches were assessed based on short InChIKey comparison.

F1 score (harmonic mean of precision and recall rate) was used in order to evaluate the impact of the *taxonomically informed scoring*. More details on the establishment of the score can be found in Material and Methods. The F1 scores of the three metabolite annotation tools before and after *taxonomically informed scoring* are displayed on Figure 4. The *taxonomically informed scoring* stage led to a systematic increase of the F1 score for the benchmarked tools. This increase was 7-fold (ISDB-DNP), 1.5-fold (Sirius) and 3-fold (MS-Finder).

Optimization of Scores Combination for the Taxonomically Informed Scoring

In order to verify our initial hypothesis and define the optimal scores combination (at the family, genus and species taxa level) to be applied for *taxonomically informed scoring* we proceeded to an optimization of the *taxonomically informed scoring* function.

To this end, the taxonomic information related to candidate annotations was artificially degraded. This step allowed to mimic a “real life” case in which candidate annotation's taxonomic metadata are not necessarily complete or correct down to the species level. Using the procedure detailed in the corresponding Material and Methods section, a Bayesian optimization algorithm was applied four times on four randomized datasets. It quickly converged (100 iterations) towards a global maximum (max 1,126 hits, see Figure 5). The optimal scores were found to be 0.81, 1.62 and 2.55 for the family, genus and species taxa level, respectively. Such scores are dependent on the nature and completeness of the employed taxonomic metadata. However, the results obtained when applying the Bayesian optimization on the annotation sets for which taxonomic metadata was randomly degraded, indicated that optimal results were systematically obtained when *the attributed scores were inversely proportional to the taxa hierarchical position*, thus confirming our initial hypothesis.

Application of the Taxonomically Informed Scoring to the Annotation of Metabolites From *Glaucium* Sp

The interest of the *taxonomically informed scoring* was further illustrated for the annotation of specialized metabolites from *Glaucium* species (Papaveraceae family). Three species, *G. grandiflorum*, *G. fimbriigerum* and *G. corniculatum* were studied. The ethyl acetate and methanolic extracts of the three species were profiled by UHPLC-HRMS in positive ionization mode using a data-dependent MS/MS acquisition. After appropriate data treatment and molecular network generation

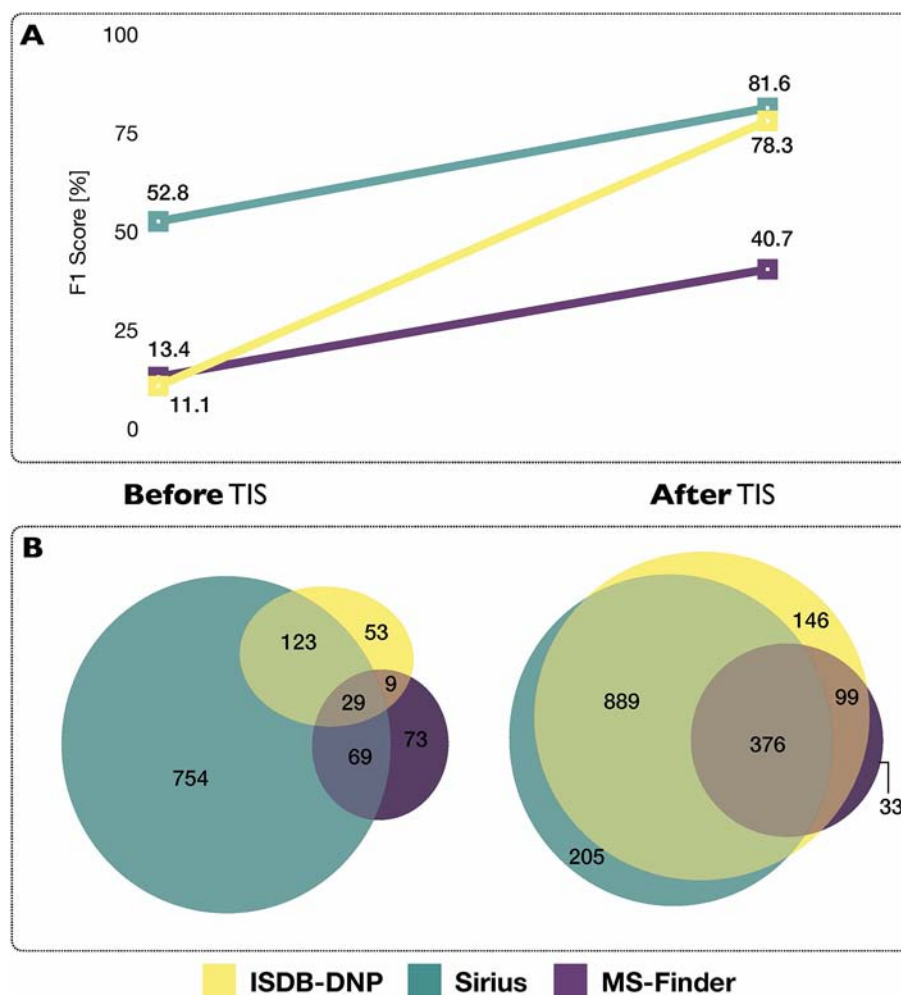


FIGURE 4 | (A) Influence of the *taxonomically informed scoring* (TIS) on the F1 score of each metabolite annotation tool. The F1 score offers a global evaluation of the precision and recall rate of the annotation process, the higher the better. **(B)** Venn diagrams representing common and unique correct annotations of each tool at rank 1 before and after the *taxonomically informed scoring* step.

(see corresponding Material and Methods section), the *taxonomically informed scoring* was used to re-rank the candidate annotation returned by the ISDB-DNP. Best five hits were kept. We especially focused on the two major compounds (MS signal intensity) of *G. grandiflorum*. These were feature m/z 342.1670 at 1.42 min and m/z 356.1860 at 1.83 min. According to the optimization results (see previous section), a score of 0.81 was given to candidates for which the biological source was found to be Papaveraceae at the family level, 1.62 to *Glaucium* at the genus level and 2.55 to *G. grandiflorum* at the species level. The results of the *taxonomically informed scoring* annotation for feature m/z 342.1670 at 1.42 min are presented in **Table 1**. See **Supplementary Table S5** for annotation results concerning feature m/z 356.1860 at 1.83 min.

Both features were targeted within the extract and, after isolation, the structure of their corresponding compound was determined by 1D and 2D NMR measurements (see spectra in **Supplementary Material S1** and **S2**). NMR spectra of feature

m/z 342.1670 at 1.42 min matched to the literature reported spectra for prednicentrine (Guinaudeau et al., 1979). NMR Spectra of feature m/z 356.1860 at 1.83 min matched to glaucine (Huang et al., 2004). In both cases, the candidate structure proposed *via* the *taxonomically informed scoring* annotation at rank 1 was found to be correct. With the classical spectral matching process, the correct candidates were initially ranked at positions 9 and 7 for prednicentrine and glaucine, respectively (see **Table 1** and **S5** in **Supplementary Material**).

Additional prednicentrine analogues were annotated in the corresponding cluster (see examples in **S6** in **Supplementary Material**).

DISCUSSION

The metabolite annotation process can be boiled down to the comparison of attributes (e.g. exact mass, molecular formula

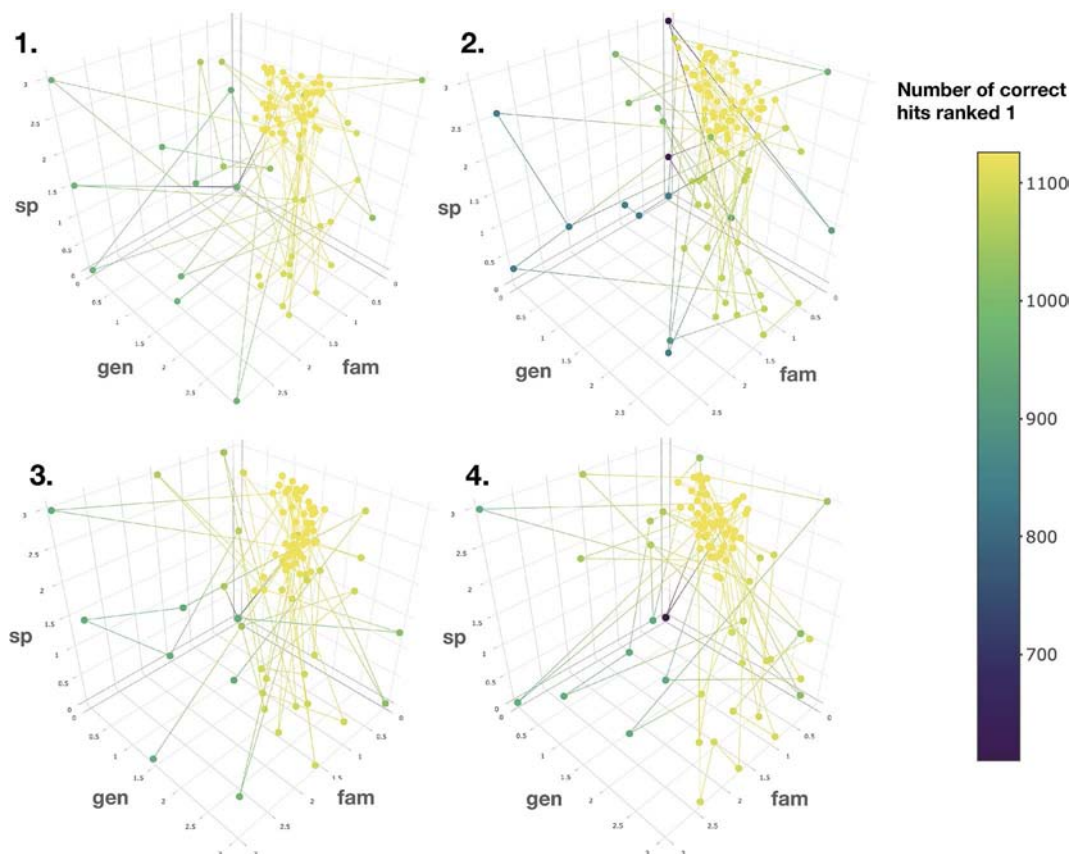



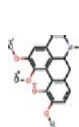
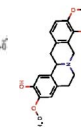
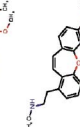
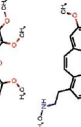
FIGURE 5 | Results of the Bayesian optimization converge toward the optimal scores combination required for a maximal number of correct annotations ranked at the first position. This is observed for four randomly degraded training sets (first optimization round displayed). sp, gen and fam axes correspond to the score given when a match is found at the species, genus or family taxa level, respectively. Results confirm that the applied scores should be *inversely proportional to the taxonomic distance between the biological source associated with the queried spectra and the biological source of the candidate structures*.

(MF) fragmentation spectra) of the queried analyte to attributes of candidate structures present in a DB. When HRMS and appropriate heuristic filters are used, the establishment of the MF of the analyte is relatively straightforward (Kind and Fiehn, 2007). However, this is not sufficient to proceed to metabolite annotation given the isomeric nature of numerous NP: over all compounds reported in the DNP, less than 10% have a unique chemical formula, the average number of compounds per molecular formula is 8.6 and a maximum of 1,274 isomers is found for $C_{15}H_{20}O_4$. With an MS1 analysis relying on exact mass only, no ranking between those isomers is possible. Computational metabolite annotation tools allow to attribute a score to candidate structures and, thus, to discriminate isomeric molecules. However, MF and fragmentation spectra are not the only attributes which can be compared in the metabolite annotation process. Specialized metabolites, as products of biosynthetic clusters themselves part of the genome, are tightly linked to the taxonomic position of the producing organisms (Hoffmann et al., 2018; Ernst et al., 2019). Here, we demonstrate that the taxonomic distance between the biological source of the queried compound and the biological source of the candidate structures is a valuable attribute to integrate into the metabolite annotation process. We show that such information can be considered in a

taxonomically informed scoring system and automatically applied to the outputs of different computational metabolite annotation programs. The consideration of taxonomic information was shown to systematically improve the F1 score of the evaluated solutions (ISDB-DNP, Sirius, MS-Finder) with a 1.5 to 7-fold increase. The advantage of considering such information in the metabolite annotation process are thus observed *independently* of the tools and their associated structural DBs.

It is worth noting that this benchmarking was carried to evaluate the importance of considering taxonomic information during the metabolite annotation process. *It was not meant to compare the performances of the tools*. Indeed, all compounds of the benchmarking dataset are present in the DNP, and the ISDB-DNP tool, which is by definition backed by the same DB is thus favored. On the other hand, the GNPS spectral libraries were also part of the Sirius training set. Furthermore filters for the selection of $[M+H]^+$ adducts and for the filtering MS/MS spectra (500 most intense peaks) were applied to meet restriction of the ISDB-DNP and MS-Finder, respectively. Finally, a number of entries (197) of the benchmarking dataset were found to have large mass difference (> 0.01 Da) between their experimental parent ion mass and their calculated exact mass. For example, cevadine $[M+H]^+$ (CCMSLIB00004689734) had an experimental

TABLE 1 | Output of the taxonomically informed scoring annotation using ISDB-DNP for feature m/z 342.1670 at 1.42 min. Predicentrine, the correct annotation, which was initially ranked at the 9th position, is ranked at the first position after taxonomically informed scoring.

Cluster ID	Structure	Short IK	Molecule Name	Family	Genus	Species	Family Score	Genus Score	Species Score	Max Taxo Score	Spectral Score	Normalized Spectral Score	Combined Spectral + Taxo Score	Rank Initial	Rank Final
1772		OUTYMWDDJURZOH	Predicentrine	Papaveraceae	Glaucium	Glaucium oxylobum	0.81	1.62	0	1.62	0.36	0.23	1.85	9	1
1772		QELDJENFQJOY	Isocondyline	Papaveraceae	Glaucium	NA	0.81	1.62	0	1.62	0.34	0.22	1.84	11	2
1772		KDFKJOFJHSVROC	Isocondyline	Papaveraceae	Glaucium	Glaucium fimbriigerum	0.81	1.62	0	1.62	0.3	0.14	1.76	28	3
1772		JADHMIUPTWPBTMT	Secosarcocapnine Me ether, N-De-Me	Papaveraceae	Sarcocapnos	Sarcocapnos crassifolia	0.81	0	0	0.81	0.4	0.32	1.13	1	4
1772		WNBUTZHPPLVTP	Secoculardine Me ether, N-de-Me	Papaveraceae	Ceratocapnos	Ceratocapnos claviculata	0.81	0	0	0.81	0.39	0.29	1.1	2	5

parent ion mass of 632.386 Da, while its calculated exact mass is 591.3407 Da (C₃₂H₄₉NO₉). Of course, such erroneous entries cannot be identified by the computational metabolite annotation tools (the list of these problematic entries is available online Problematic_entries.csv). Altogether these elements prevent a fair comparison of each tool's performances. Another precautionary statement concerns the results of the optimization on candidate datasets for which taxonomic information had been randomly degraded at multiple taxa level. This optimization indicated, for the ISDB-DNP results, that the optimal combination of scores was 0.81, 1.62 and 2.55 (family, genus and species taxa level, respectively). Such results should be taken with caution, and not as absolute optimal values, as such optimization process are heavily dependent on the training sets. Nevertheless, the optimization indicates that the best results were repeatedly obtained when the assigned scores were inversely proportional to the taxonomic distance between the biological sources of, both, the queried spectra and the candidate structures.

Other limitations of the described metabolite annotation strategy include its application range and prerequisites. Indeed, it is important to note that such *taxonomically informed scoring* system will mostly benefit the annotation process of *specialized metabolites* and not ubiquitous molecules (e.g. coming from the primary metabolism) for obvious reasons. Furthermore, it heavily depends on the availability and quality of DBs compiling structures and their biological sources reported as a fully and homogeneously resolved taxonomy. To the best of our knowledge, such DBs are not publicly available and downloadable at the moment. KNApSACk (<http://www.knapsackfamily.com/KNApSACk/>) is a comprehensive species-metabolite relationship database compiling 116,315 metabolite-species pairs entries, it is accessible online but not downloadable. Other databases such as FooDB (<http://foodb.ca>) are fully downloadable but however focused on food-related metabolites, furthermore the biological sources field is not standardized. The NPAtlas (<https://www.npatlas.org/>) is an interesting initiative, however biological sources information down to the species level is only accessible in query mode and the DB is limited to 24,594 metabolites of microbial origin only. The Dictionary of Natural Products, which we used in this study is the widest compilation of structure/biological sources pairs, but is only available commercially. Furthermore, the biological sources are reported as a free text field (codes are available only for the family taxa levels and above), thus requiring tedious standardization and name resolving.

It is therefore important for the community to start the systematic reporting of biological sources, together with spectral and structural information, when documenting novel metabolites. In fact, reporting newly described biological occurrence should be encouraged even for previously described metabolites. However, the policy of most journals in NP research is to accept for publication only description of novel and bioactive structures, which hinders these potentially informative reports. The GNPS spectral libraries (<https://gnps.ucsd.edu/ProteoSAFe/libraries.jsp>) and MassIVE repositories (<https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp>) appear as optimal places, at the moment, to compile and share NP spectral and structural information. However, although

free text comments can complement the documentation of an entry, no standardized fields are available to report the biological sources of the uploaded spectra. The creation of such a feature, ideally directly linking the entered biological sources to existing taxonomy backbones such as GBIF (<https://www.gbif.org>) or Catalogue of Life (<http://www.catalogueoflife.org>), would be extremely useful. A recent initiative, the Pharmacognosy Ontology (PHO), that builds on 50 years of development of NAPRALERT (<https://www.napralert.org>) is aimed at providing a Free and Open resource that will link taxonomical, chemical and biological data (http://ceur-ws.org/Vol-1747/IP12_ICBO2016.pdf). Of course, and in addition to correct and systematic biological sources occurrence reporting in dedicated DBs, it is of utmost importance to count on the expert knowledge of trained taxonomists specialized in the classification of living organisms. But it seems that today, unfortunately, these people are very few (Ajmal Ali and Choudhary, 2011; Drew, 2011).

Building on the proposed *taxonomically informed scoring*, further developments will pass by the consideration of more accurate quantification of taxonomic distances and by strengthening the metascoring system. Indeed, the approach presented here only considers the identity between the biological sources, at different taxa level, of the query compounds and the ones of the candidate structures. Taking into account a more precise phylogenetic position within or across taxa, for example *via* the calculation of taxonomic distinctiveness indexes (Clarke and Warwick, 1998; Weikard et al., 2006), could offer a more accurate distance and eventually improve such *taxonomically informed scoring* process. Such calculations could however reveal to be computationally demanding to realize on the fly. On another plan, efforts remain to be done towards the establishment of a global metascore (see **Figure 1**). Integrating the proposed taxonomic distance scoring (S_2 in **Figure 1**) with the maximal number of available metadata (S_3, S_4, \dots) when proceeding to metabolite annotation should only be beneficial to such process. However, problematics such as the individual weights (w_1, w_2, w_3, \dots) to attribute to each individual score of the metascore will have to be addressed.

CONCLUSION

Efficient characterization of specialized metabolomes is a key challenge in metabolomics and NP research. Recent technical advances allow access to an ever-increasing amount of data, raising the need for *ad hoc* computational solutions for their interpretation. The metabolite annotation process, which can be resumed to the comparison of attributes of the queried features against attributes of the candidate structures can benefit from information complementary to the classically used MS/MS fragmentation. Ideally, the quantification of multiple attributes' similarities (or dissimilarities) should be integrated within a metascoring system. Here, we demonstrate that the consideration of the taxonomic distance separating the biological sources of both the queried analytes and the candidate structures can drastically improve the efficiency of existing MS-based computational metabolite annotation solutions. Metabolite annotation is crucial to guide chemical ecology research or drug discovery projects. More than

two hundred years later, the present work thus supports the first of De Candolle's assumptions, "*Plant taxonomy would be the most useful guide to man in his search for new industrial and medicinal plants*" (de Candolle, 1804). His correlated postulate, "*Chemical characteristics of plants will be most valuable to plant taxonomy in the future*", will be equally interesting to verify with computational approaches. Various strategies have been proposed to exploit structural (or biosynthetic) relationships among metabolites and further organize the producing organisms (Liu et al., 2017; Junker, 2018; Ernst et al., 2019; Kang et al., 2019) and interesting developments will appear once robust metabolite annotation solution are coupled to comprehensive DBs compiling structures and their biological sources. Indeed, specialized metabolome annotation could be a novel way to infer the taxonomic position of an unknown sample, just as valid as a genetic sequencing. Metabolite annotation can benefit from taxonomy and taxonomic relationships can be inferred from precise metabolite characterization. Efforts in both directions should thus fuel a *virtuous cycle of research aiming to better understand Life and its chemistry*.

MATERIAL AND METHODS

Outline and Implementation of the Taxonomically Informed Scoring System

To evaluate the importance of considering taxonomic information in the annotation process, three different computational mass spectrometry-based metabolite annotation tools were used (namely, ISDB-DNP, MS-Finder and Sirius). This resulted in three different outputs constituted by a list of candidates returned by each tool for the entries of the benchmarking dataset. These candidates were ranked according to the scoring system of each tool. R scripts in the form of markdown notebooks were written to perform 1) cleaning and standardization of the outputs (1_taxo_cleaner.Rmd) 2) *taxonomically informed scoring* and re-ranking (2_taxo_scorer.Rmd). First, the outputs were standardized to a table containing on each row: a unique spectral identifier (CCMSLIB N°) of the queried spectra, the short InChIKey of the candidate structures, the score of the candidates (within the scoring system of the used metabolite annotation tool), the biological source of the standard compound and the biological source of the candidate structures. As described in the results section, a score, inversely proportional to the taxonomic distance between the biological source of the annotated compound and the one biological source of the candidate structure, was given when an exact match was found between both biological sources at the family, genus or/and species level(s). A sum of this score and the original score yielded the *taxonomically informed score*. This score was then used to re-rank the candidates. See **Figure 2** for a schematic overview of the *taxonomically informed scoring* process. Scripts are available at https://github.com/oolonek/taxo_scorer.

Dataset Preparation

Structural and Biological Sources Dataset

In the Dictionary of Natural Products (v 27.1), taxonomic information appears in two fields. The *Biological Source* field, which

is constituted by a free text field reporting occurrence of a specific compound and the *Compound Type* field which reports various codes corresponding to molecule classes or taxonomic position at the family level. As an example, for the entry corresponding to larictritrin 3-glucoside (ODXINVINFDDDD-UHFFFAOYSA-N), the Biological Source field indicates “Isol. from *Larix* spp., *Cedrus* sp. and other plant spp. Constit. of *Vitis vinifera* cv. Petit Verdot grapes and *Abies amabilis*,” and the Compound Type field indicates “V.K.52600 W.I.40000 W.I.35000 Z.N.50000 Z.Q.71600” suggesting that biological sources are found in the Phyllocladaceae (Z.N.50000) and Vitaceae family (Z.Q.71600). The biological source information is reported in a non-homogeneous way and multiple biological sources are reported in the same row. In order to extract taxonomic information out of the free text contents, the *gnfinder* program (<https://github.com/gnames/gnfinder>) was used. Gnfinder takes UTF8-encoded text as inputs and returns back JSON-formatted output that contains detected scientific names. It automatically detects the language of the text and uses complementary heuristic and natural language processing algorithms to detect patterns corresponding to scientific binomial or uninomial denomination. Gnfinder was forced for English language detection. In addition to scientific denomination extraction, gnfinder allows to match the detected names against the Global Names index services (<https://index.globalnames.org>). The preferred taxonomy backbone was set to be Catalogue of Life. This last step allowed to return the full taxonomy down to the entered taxa level. It also allows to resolve synonymy. Since gnfinder is designed to mine raw texts, the JSON formatted output indicates the position of the detected name in the original input by character position. A python script was written to output a .csv file with the found name and taxonomy in front of the corresponding input. When multiple biosources were found for an entry, this one was duplicated in order to obtain a unique structure/biological source pair per row. The script is available online ([gnfinder_field_scrapper.py](#)).

Structural and Spectral Dataset

All GNPS libraries and publicly accessible third-party libraries were retrieved online (<https://gnps.ucsd.edu/ProteoSAFe/libraries.jsp>) and concatenated as a single spectral file (Full_GNPS_lib.mgf) in the .mgf format. A python Jupyter notebook ([mgf_filterer.ipynb](#)) was created to filter.mgf spectral file according to specific parameters: maximum and minimum number of fragments per spectrum and defined spectral ID (e.g. CCMLIB N°). The spectral file was filtered to retain only entries having at least 6 fragments. For spectra containing more than 500 fragments, only the 500 most intense were kept. A second python Jupyter notebook (GNPS_parser_cleaner.ipynb) was written to proceed to 1) extraction of relevant metadata (parent ion mass, SMILES, InChI, library origin, source instrument, molecule name and individual spectrum id value (CCMSLIB N°) 2) filtering entries having at least one structural information associated (SMILES and/or InChI) and corresponding to protonated adducts and 3) converting structures to their InChIKey, a 27-character hashed version of the full InChI. The InChIKey conversion was realized using the RDKit 2019.03.1 framework (RDKit: Open-source cheminformatics; <http://www.rdkit.org>). This resulted in a structural dataset (GNPS_lib_structural.tsv) of 40138 entries

constituted by 8558 unique compounds. The dataset was further filtered to keep entries which parent masses were comprised between 100 and 1,500 Da. Duplicate structures and stereoisomers were removed by keeping distinct InChIKey according to the first layer (first 14 characters) of the hash code. This spectral dataset encompasses spectra acquired on a variety of MS platforms (See **Supplementary Figure S7**). Scripts are available at https://github.com/oologonek/taxo_scorer. Input and output data are available on OSF at the following address (<https://osf.io/bvs6x/>).

Structural, Spectral, and Biological Sources Dataset (Benchmarking Dataset)

Once the structural and biological sources dataset and the structural and spectral datasets were prepared (as described above), both were joined in order to attribute a biological source to each spectrum. The scripts used to proceed to the merging step are part of the python Jupyter notebook (GNPS_parser_cleaner.ipynb). Since in most cases it is not expected to differentiate stereoisomers based on their MS spectra, the combination of both datasets was made using the short InChIKey (first 14 characters of the InChIKey) as a common key. In this merging process, only entries having biological source information resolved against the Catalogue of Life and complete down to the species level were retained. However, this merging implied that, for a given biological source, the information on the 3D aspects of the structure was lost. While this was not an issue for the benchmarking objective of this work the resulting dataset does not constitute a reliable occurrence dataset for annotation that needs stereoisomers to be differentiated. The resulting dataset containing structural, spectral and biological sources information was constituted by 2107 distinct entries and was named *benchmarking dataset*. The scripts allowing to generate the benchmarking dataset are available at https://github.com/oologonek/taxo_scorer. The benchmarking dataset spectral data (benchmarking_dataset_spectral.mgf), and associated metadata (benchmarking_dataset_metadata.tsv) are available at the following address (<https://osf.io/bvs6x/>).

Computational Metabolite Annotation Tools

ISDB-DNP

The ISDB-DNP (*In Silico* DataBase—Dictionary of Natural Products) is a metabolite annotation workflow that we previously developed (Allard et al., 2016). A version using the freely available Universal Natural Products Database (ISDB-UNPD) is available online (<http://oologonek.github.io/ISDB/>). This approach is focused on specialized metabolites annotation and is constituted by a pre-fragmented theoretical spectral DB version of the DNP. The *in silico* fragmentation was performed using CFM-ID, a software using a probabilistic generative model for the fragmentation process, and a machine learning approach for learning parameters for this model from MS/MS data (Allen et al., 2015). CFM, is, to the best of our knowledge, the only solution available at the moment allowing to output a spectrum with fragment intensity prediction. The matching phase between experimental spectra and the theoretical DB is based on a spectral similarity computation performed using Tremolo as a spectral

library search tool (Wang and Bandeira, 2013). The parameters used to proceed to the benchmarking dataset analysis were the following: parent mass tolerance 0.05 Da, minimum cosine score 0.1, no limits for the number of returned candidates.

MS-Finder

This *in silico* fragmentation approach considers multiple parameters such as bond dissociation energies, mass accuracies, fragment linkages and various hydrogen rearrangement rules at the candidate ranking phase (Tsugawa et al., 2016). The resulting scoring system range from 1 to 10. The parameters used to proceed to the benchmarking dataset analysis were the following: mass tolerance setting: 0.1 Da (MS1), 0.1 Da (MS2); relative abundance cut off: 5% formula finder settings: LEWIS and SENIOR check (yes), isotopic ratio tolerance: 20%, element probability check (yes), element selection (O, N, P, S, Cl, Br). Structure Finder setting: tree depth: 2, maximum reported number: 100, data sources (all except MINEs DBs. Total number of structures, 321,617.) MS-Finder v. 3.22 was used, it is available at the following address: http://prime.psc.riken.jp/Metabolomics_Software/MS-FINDER/.

Sirius

Sirius 4.0.1 is considered as a state-of-the-art metabolite annotation solution, which combines molecular formula calculation and the prediction of a molecular fingerprint of a query compound from its fragmentation tree and spectrum (Dührkop et al., 2019). Sirius uses a DB of 73,444,774 unique structures for its annotations. The parameters used to proceed to the benchmarking dataset analysis were the following for Sirius molecular formula calculation: possible ionization [M+H]⁺, instrument: Q-TOF, ppm tolerance 50 ppm, Top molecular formula candidates: 3, filter: formulas from biological DBs. For the CSI : FingerID step, the parameters were the following: possible adducts: [M+H]⁺, filter: compounds present in biological DB, maximal number of returned candidates: unlimited. Sirius 4.0.1 is available at the following address: <https://bio.informatik.uni-jena.de/software/sirius/>.

Results Analysis

The F1 score was calculated for each evaluated metabolite annotation tool before and after the *taxonomically informed scoring* step. The F1 score is the harmonic mean of the recall (True Positive/(True Positive + False Negative)) and precision rate (True Positive/(True Positive + False Positive)) of a tool. The True Positive (TP) corresponds to the number of correct candidate annotations at rank 1, the False Positive (FP) to the number of wrong candidate annotations at rank 1, and the False Negative (FN) to the number of correct annotations at rank >1. The F1 score is then calculated as follows:

$$F1\ score = 2 \times \frac{(Recall\ rate \times Precision\ rate)}{(Recall\ rate + Precision\ rate)}$$

An R notebook to analyze the results of the *taxonomically informed scoring* process and plot the figures of this manuscript is available online (taxo_figures.Rmd) at https://github.com/oolonek/taxo_scorer.

Optimization of the Scores Combination for the Taxonomically Informed Scoring

In order to establish the optimal scores to be applied for each of the taxonomic distances (family, genus and species), the information related to candidate annotations was artificially degraded. For this, the annotation dataset returned by the ISDB-DNP approach against the benchmarking dataset was randomized. The randomized annotation dataset was then split into four equal blocks. For the first three blocks, the biological source information was deleted, respectively, at the species level; at the genus and species level; and, finally, at the family, genus and species levels. The fourth block was not modified. Finally, the four blocks were merged back to a unique dataset. The process was repeated four times yielding four datasets with randomly incomplete biological sources. The taxonomic distance informed scoring process was compiled to a unique function taking three arguments (scores given when a match was found at the family, genus and species level, respectively) and outputting the number of correct hits ranked at the first position. A parallelizable Bayesian optimization algorithm (<https://github.com/AnotherSamWilson/ParBayesianOptimization>) was then used, being particularly suited for the optimization of black box functions for which no formal representation is available (arXiv:1807.02811). The bounds were set between 0 and 3 for the exploration of the three parameters of the function. Number of initial points was set to 10 and the number of iterations to 100. Parameter kappa (κ) was set to 5.152, to force the algorithm to explore unknown areas. The chosen acquisition function was set to Expected Improvement (ei). Epsilon parameter (ε, eps) was set to 0. The whole procedure was run 4 times on the 4 randomized datasets. Best set of parameters were then averaged across the 16 results set. All codes required for this optimization step are available online (3_taxo_optimizer.Rmd) at https://github.com/oolonek/taxo_scorer.

Chemical Analysis and Isolation of Compounds From Glaucium Extract Plant Material

The aerial flowering parts of three *Glaucium* species were collected in May and June of 2015 from the northern part of Iran including Mazandaran and Tehran provinces. The samples were identified by Dr. Ali Sonboli, Medicinal Plants and Drugs Research Institute, Shahid Beheshti University, Tehran, Iran. The voucher specimens MPH-2351 for *G. grandiflorum* (vernacular Shaghayegh goldrosht), MPH-2352 for *G. fimbriigerum* (vernacular Shaghayegh sharabeie) and MPH-2353 for *G. corniculatum* (vernacular Shaghayegh shakhdar or red horned poppy) have been deposited at the Herbarium of Medicinal Plants and Drugs Research Institute (HMPDRI), Shahid Beheshti University, Tehran, Iran.

Mass Spectrometry Analysis

Chromatographic separation was performed on a Waters Acquity UPLC system interfaced to a Q-Exactive Focus mass spectrometer (Thermo Scientific, Bremen, Germany), using a heated electrospray ionization (HESI-II) source. Thermo Scientific Xcalibur 3.1 software was used for instrument control. The LC conditions were as follows: column, Waters BEH C18 50 × 2.1 mm, 1.7 μm; mobile

phase, (A) water with 0.1% formic acid; (B) acetonitrile with 0.1% formic acid; flow rate, 600 $\mu\text{L}\cdot\text{min}^{-1}$; injection volume, 6 μL ; gradient, linear gradient of 5–100% B over 7 min and isocratic at 100% B for 1 min. The optimized HESI-II parameters were as follows: source voltage, 3.5 kV (pos); sheath gas flow rate (N_2), 55 units; auxiliary gas flow rate, 15 units; spare gas flow rate, 3.0; capillary temperature, 350.00°C, S-Lens RF Level, 45. The mass analyzer was calibrated using a mixture of caffeine, methionine–arginine–phenylalanine–alanine–acetate (MRFA), sodium dodecyl sulfate, sodium taurocholate, and Ultramark 1621 in an acetonitrile/methanol/water solution containing 1% formic acid by direct injection. The data-dependent MS/MS events were performed on the three most intense ions detected in full scan MS (Top3 experiment). The MS/MS isolation window width was 1 Da, and the stepped normalized collision energy (NCE) was set to 15, 30 and 45 units. In data-dependent MS/MS experiments, full scans were acquired at a resolution of 35,000 FWHM (at m/z 200) and MS/MS scans at 17,500 FWHM both with an automatically determined maximum injection time. After being acquired in a MS/MS scan, parent ions were placed in a dynamic exclusion list for 2.0 s.

MS Data Pretreatment

The MS data were converted from .RAW (Thermo) standard data format to .mzXML format using the MSConvert software, part of the ProteoWizard package (Chambers et al., 2012). The converted files were treated using the MZMine software suite v. 2.38 (Pluskal et al., 2010).

The parameters were adjusted as following: the centroid mass detector was used for mass detection with the noise level set to 1.0E6 for MS level set to 1, and to 0 for MS level set to 2. The ADAP chromatogram builder was used and set to a minimum group size of scans of 5, minimum group intensity threshold of 1.0E5, minimum highest intensity of 1.0E5 and m/z tolerance of 8.0 ppm. For chromatogram deconvolution, the algorithm used was the wavelets (ADAP). The intensity window S/N was used as S/N estimator with a signal to noise ratio set at 25, a minimum feature height at 10,000, a coefficient area threshold at 100, a peak duration ranges from 0.02 to 0.9 min and the RT wavelet range from 0.02 to 0.05 min. Isotopes were detected using the isotopes peaks grouper with a m/z tolerance of 5.0 ppm, a RT tolerance of 0.02 min (absolute), the maximum charge set at 2 and the representative isotope used was the most intense. An adduct (Na^+ , K^+ , NH_4^+ , CH_3CN^+ , CH_3OH^+ , $\text{C}_3\text{H}_8\text{O}^+$ (IPA $^+$)) search was performed with the RT tolerance set at 0.1 min and the maximum relative peak height at 500%. A complex search was also performed using $[\text{M}+\text{H}]^+$ for ESI positive mode, with the RT tolerance set at 0.1 min and the maximum relative peak height at 500%. Peak alignment was performed using the join aligner method (m/z tolerance at 8 ppm), absolute RT tolerance 0.065 min, weight for m/z at 10 and weight for RT at 10. The peak list was gap-filled with the same RT and m/z range gap filler (m/z tolerance at 8 ppm). Eventually the resulting aligned peaklist was filtered using the peak-list rows filter option in order to keep only features associated with MS2 scans.

Molecular Networks Generation

In order to keep the retention time, the exact mass information and to allow for the separation of isomers, a feature based molecular

network (<https://ccms-ucsd.github.io/GNPSDocumentation/featurebasedmolecularnetworking/>) was created using the .mgf file resulting from the MZMine pretreatment step detailed above. Spectral data was uploaded on the GNPS molecular networking platform. A network was then created where edges were filtered to have a cosine score above 0.7 and more than six matched peaks. Further edges between two nodes were kept in the network if and only if each of the nodes appeared in each other's respective top 10 most similar nodes. The spectra in the network were then searched against GNPS' spectral libraries. All matches kept between network spectra and library spectra were required to have a score above 0.7 and at least six matched peaks. The output was visualised using Cytoscape 3.6 software (Shannon et al., 2003). The GNPS job parameters and resulting data are available at the following address (<https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=a475a78d9ae8484b904bcd7a16abd1f>).

Taxonomically Informed Metabolite Annotation

The spectral file (.mgf) and attributes metadata (.clustersummary) obtained after the MN step were annotated using the ISDB-DNP with the following parameters: parent mass tolerance 0.005 Da, minimum cosine score 0.2, maximal number of returned candidates: 50. An R script was written to proceed to the *taxonomically informed scoring* on GNPS outputs and return an attribute table which can be directly loaded in Cytoscape. The script is available online (taxo_scorer_user.Rmd) at https://github.com/ooolonek/taxo_scorer.

Isolation of Predicentrine and Glaucine From *G. grandiflorum*

The air-dried, ground and powdered plant materials (500 g) was successively extracted by solvents of increasing polarities (hexane, ethyl acetate and methanol), 4 \times 5.0 L of each solvent (48 h). An aliquot of each ethyl acetate and methanolic extract was submitted to C18 SPE (eluted with 100% MeOH), dried under nitrogen flow and redissolved at 5 mg/ml in MeOH for LC–MS analysis. The methanolic extract of *G. grandiflorum* was concentrated under reduced pressure, then dried with a nitrogen flow until complete evaporation of the residual solvent yielding 50 g of extract. An aliquot (5 g) was subjected to a VLC in order to eliminate sugars and other very polar compounds. A 250 ml sintered-glass Buchner funnel connected to a vacuum line was packed with a C18 reverse phase LiChroprep 40–63 μm (Lobar Merck, Darmstadt, Germany). After conditioning the stationary phase with methanol (4 \times 250 ml, 0.1% formic acid) and distilled water (4 \times 250 ml, 0.1% formic acid), 5 g of methanolic extract was dissolved in water and the mixture was deposited on the stationary phase. Elution of the sample was conducted using water (4 \times 250 ml, 0.1% formic acid) in the first step and followed by methanol (4 \times 250 ml, 0.1% formic acid) in the second step. This process yielded 1.4 g of processed methanolic extract. After condition optimisation at the analytical level, 50 mg of the extract were solubilized in 500 μL DMSO and injected using a Rheodyne® valve (1 ml loop). Semi-preparative HPLC–UV purification was performed on a Shimadzu system equipped with: LC20A module elution pumps, an SPD-20A UV/VIS detector, a 7725I Rheodyne® injection valve, and a FRC-10A fraction collector (Shimadzu, Kyoto, Japan). The HPLC system was controlled by

the LabSolutions software. The HPLC conditions were selected as follows: Waters X-Bridge C18 column (250 × 19 mm i.d., 5 μm) equipped with a Waters C18 pre-column cartridge holder (10 × 19 mm i.d.); solvent system consists of ACN (2 mM TEA) (B) and H₂O (2 mM TEA & 2 mM ammonium acetate) (A). Optimized separation condition from the analytical was transferred to semi-preparative scale by a geometric gradient transfer software (Guillarme et al., 2008). The separation was conducted in gradient elution mode as follows: 5% B in 0–5 min, 12% B in 5–10 min, 30% B in 10–30 min, 60% B in 30–55 min, 100% B in 55–65 min. The column was reconditioned by equilibration with 5% of B in 15 min. Flow rate was equal to 17 ml/min and UV traces were recorded at 210 nm and 280 nm. The separation procedure yielded 0.3 mg of predicitrine and 3.4 mg of glaucine. Spectra for predicitrine (CCMSLIB00005436122) and glaucine (CCMSLIB00005436123) were deposited on GNPS servers.

NMR Analysis

The NMR spectra of each isolated compound was recorded on a Bruker BioSpin 600 MHz spectrometer (Avance Neo 600). Chemical shifts (δ) were recorded in parts per million in methanol-d₄ with TMS as an internal standard. NMR data are available as **Supplementary Material S1** and **S2**.

DATA AVAILABILITY STATEMENT

Scripts and datasets generated and analyzed for this study can be found on github (scripts): https://github.com/oolonek/taxo_scorer and at the following OSF repository (datasets): [https://osf.io/bvs6x/\(DOI 10.17605/OSF.IO/BVS6X\)](https://osf.io/bvs6x/(DOI%2010.17605/OSF.IO/BVS6X)). This manuscript has been released as a pre-print at bioRxiv: <http://dx.doi.org/10.1101/702308>.

REFERENCES

- Ajmal Ali, M., and Choudhary, R. K. (2011). India needs more plant taxonomists. *Nature* 471, 37. doi: 10.1038/471037d
- Allard, P.-M., Genta-Jouve, G., and Wolfender, J.-L. (2017). Deep metabolome annotation in natural products research: towards a virtuous cycle in metabolite identification. *Curr. Opin. Chem. Biol.* 36, 40–49. doi: 10.1016/j.cbpa.2016.12.022
- Allard, P.-M., Péresse, T., Bisson, J., Gindro, K., Marcourt, L., Pham, V. C., et al. (2016). Integration of molecular networking and in-silico ms/ms fragmentation for natural products dereplication. *Anal. Chem.* 88, 3317–3323. doi: 10.1021/acs.analchem.5b04804
- Allen, F., Greiner, R., and Wishart, D. (2015). Competitive fragmentation modeling of ESI-MS/MS spectra for putative metabolite identification. *Metabolomics* 11, 98–110. doi: 10.1007/s11306-014-0676-4
- Allen, F., Pon, A., Wilson, M., Greiner, R., and Wishart, D. (2014). CFM-ID: a web server for annotation, spectrum prediction and metabolite identification from tandem mass spectra. *Nucleic Acids Res.* 42, W94–W99. doi: 10.1093/nar/gku436
- Bach, E., Szedmak, S., Brouard, C., Böcker, S., and Rousu, J. (2018). Liquid-chromatography retention order prediction for metabolite identification. *Bioinformatics* 34, i875–i883. doi: 10.1093/bioinformatics/bty590
- Bauer, C. A., and Grimme, S. (2016). How to compute electron ionization mass spectra from first principles. *J. Phys. Chem. A* 120, 3755–3766. doi: 10.1021/acs.jpca.6b02907
- Böcker, S., Letzel, M. C., Lipták, Z., and Pervukhin, A. (2009). SIRIUS: decomposing isotope patterns for metabolite identification. *Bioinformatics* 25, 218–224. doi: 10.1093/bioinformatics/btn603

AUTHOR CONTRIBUTIONS

P-MA, AR, MD-K, and J-LW designed the study. JB wrote the Python script for gfinder output formatting. P-MA and AR wrote the scripts for dataset preparation, *taxonomically informed scoring*, and results analysis. P-MA, AR, MD-K, SO, MB, and TS used the scripts for metabolite annotation and provided feedback. MB and SNE collected the *Glaucium* species. P-MA and AR performed the LCMS analysis on the *Glaucium* extracts. MB analyzed profiling data, isolated the compounds of *Glaucium*, and established their structures. P-MA wrote the manuscript together with AR and J-LW. All authors discussed the results and commented on the manuscript.

FUNDING

JB gratefully acknowledges the support of this work by grant U41 AT008706 from NCCIH and ODS. J-LW is thankful to the Swiss National Science Foundation for the support in the acquisition of the NMR 600 MHz (SNF R'Equip grant 316030_164095).

ACKNOWLEDGMENTS

MD-K acknowledges gratefully the support by the Yamada Science Foundation and The Nagai Foundation Tokyo. The authors acknowledge Ali Bakiri for fruitful discussions on the optimization of the weights.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.01329/full#supplementary-material>

- Brunetti, A. E., Carnevale Neto, F., Vera, M. C., Taboada, C., Pavarini, D. P., Bauermeister, A., et al. (2018). An integrative omics perspective for the analysis of chemical signals in ecological interactions. *Chem. Soc. Rev.* 47, 1574–1591. doi: 10.1039/C7CS00368D
- Chambers, M. C., Maclean, B., Burke, R., Amodei, D., Ruderman, D. L., Neumann, S., et al. (2012). A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* 30, 918–920. doi: 10.1038/nbt.2377
- Clarke, K. R., and Warwick, R. M. (1998). A taxonomic distinctness index and its statistical properties. *J. Appl. Ecol.* 35, 523–531. doi: 10.1046/j.1365-2664.1998.3540523.x
- da Silva, R. R., Wang, M., Nothias, L.-F., van der Hooft, J. J. J., Caraballo-Rodríguez, A. M., Fox, E., et al. (2018). Propagating annotations of molecular networks using in silico fragmentation. *PLoS Comput. Biol.* 14, e1006089. doi: 10.1371/journal.pcbi.1006089
- de Candolle, A. P. (1804). *Essai sur les propriétés médicales des plantes comparées avec leurs formes extérieures et leur classification naturelle*. Available at: <https://gallica.bnf.fr/ark:/12148/bpt6k8511237>
- Djombou-Feunang, Y., Pon, A., Karu, N., Zheng, J., Li, C., Arndt, D., et al. (2019). CFM-ID 3.0: Significantly improved ESI-MS/MS prediction and compound identification. *Metabolites* 9, 72. doi: 10.3390/metabo9040072
- Drew, L. W. (2011). Are we losing the science of taxonomy? *BioScience* 61, 942–946. doi: 10.1525/bio.2011.61.12.4
- Dührkop, K., Fleischauer, M., Ludwig, M., Aksenov, A. A., Melnik, A. V., Meusel, M., et al. (2019). SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nat. Methods* 16, 299–302. doi: 10.1038/s41592-019-0344-8

- Ernst, M., Nothias, L.-F., van der Hooft, J. J. J., Silva, R. R., Saslis-Lagoudakis, C. H., Grace, O. M., et al. (2019). Assessing specialized metabolite diversity in the cosmopolitan plant genus *euphorbia* L. *Front. Plant Sci.* 10, 846. doi: 10.3389/fpls.2019.00846
- Gaudêncio, S. P., and Pereira, F. (2015). Dereplication: racing to speed up the natural products discovery process. *Nat. Prod. Rep.* 32, 779–810. doi: 10.1039/C4NP00134F
- Guillarme, D., Nguyen, D. T. T., Rudaz, S., and Veuthey, J.-L. (2008). Method transfer for fast liquid chromatography in pharmaceutical analysis: application to short columns packed with small particle. Part II: gradient experiments. *Eur. J. Pharmaceut. Biopharmaceut.* 68, 430–440. doi: 10.1016/j.ejpb.2007.06.018
- Guinaudeau, H., Leboeuf, M., and Cavé, A. (1979). Aporphine alkaloids. II. *J. Nat. Prod.* 42, 325–360. doi: 10.1021/np50004a001
- Hoffmann, T., Krug, D., Bozkurt, N., Duddela, S., Jansen, R., Garcia, R., et al. (2018). Correlating chemical diversity with taxonomic distance for discovery of natural products in myxobacteria. *Nat. Commun.* 9, 803. doi: 10.1038/s41467-018-03184-1
- Huang, W.-J., Singh, O. V., Chen, C.-H., and Lee, S.-S. (2004). Synthesis of (.-)-Glaucine and (.-)-Neospirodienone via a one-pot bischler–napieralski reaction and oxidative coupling by a hypervalent iodine reagent. *Helv. Chim. Acta* 87 (1), 167–174. doi: 10.1002/hlca.200490005
- Junker, R. R. (2018). A biosynthetically informed distance measure to compare secondary metabolite profiles. *Chemoecology* 28, 29–37. doi: 10.1007/s00049-017-0250-4
- Kang, K. B., Ernst, M., Hooft, J. J. J., Silva, R. R., Park, J., Medema, M. H., et al. (2019). Comprehensive mass spectrometry-guided phenotyping of plant specialized metabolites reveals metabolic diversity in the cosmopolitan plant family Rhamnaceae. *Plant J.* 98, 1134–1144. doi: 10.1111/tpj.14292
- Kind, T., and Fiehn, O. (2007). Seven golden rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinformatics* 8, 105. doi: 10.1186/1471-2105-8-105
- Kind, T., Tsugawa, H., Cajka, T., Ma, Y., Lai, Z., Mehta, S. S., et al. (2018). Identification of small molecules using accurate mass MS/MS search. *Mass Spectrom. Rev.* 37, 513–532. doi: 10.1002/mas.21535
- Liu, K., Abdullah, A. A., Huang, M., Nishioka, T., Altaf-Ul-Amin, M., and Kanaya, S. (2017). Novel Approach to classify plants based on metabolite-content similarity. *Biomed. Res. Int.* 2017, 5296729. doi: 10.1155/2017/5296729
- Pichersky, E., and Lewinsohn, E. (2011). Convergent evolution in plant specialized metabolism. *Annu. Rev. Plant Biol.* 62, 549–566. doi: 10.1146/annurev-arplant-042110-103814
- Pluskal, T., Castillo, S., Villar-Briones, A., and Orešič, M. (2010). MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* 11, 395. doi: 10.1186/1471-2105-11-395
- Rutties, C., Schymanski, E. L., Wolf, S., Hollender, J., and Neumann, S. (2016). MetFrag relaunched: incorporating strategies beyond in silico fragmentation. *J. Cheminform.* 8, 3. doi: 10.1186/s13321-016-0115-9
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Shen, B. (2015). A new golden age of natural products drug discovery. *Cell* 163, 1297–1300. doi: 10.1016/j.cell.2015.11.031
- Tsugawa, H., Kind, T., Nakabayashi, R., Yukihiro, D., Tanaka, W., Cajka, T., et al. (2016). Hydrogen rearrangement rules: computational MS/MS fragmentation and structure elucidation using MS-FINDER software. *Anal. Chem.* 88, 7946–7958. doi: 10.1021/acs.analchem.6b00770
- Tsugawa, H., Nakabayashi, R., Mori, T., Yamada, Y., Takahashi, M., Rai, A., et al. (2019). A cheminformatics approach to characterize metabolomes in stable-isotope-labeled organisms. *Nat. Methods* 16, 295–298. doi: 10.1038/s41592-019-0358-2
- Wang, M., and Bandeira, N. (2013). Spectral library generating function for assessing spectrum-spectrum match significance. *J. Proteome Res.* 12, 3944–3951. doi: 10.1021/pr400230p
- Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., et al. (2016). Sharing and community curation of mass spectrometry data with global natural products social molecular networking. *Nat. Biotechnol.* 34, 828–837. doi: 10.1038/nbt.3597
- Weikard, H.-P., Punt, M., and Wesseler, J. (2006). Diversity measurement combining relative abundances and taxonomic distinctiveness of species. *Divers. Distrib.* 12, 215–217. doi: 10.1111/j.1366-9516.2006.00234.x
- Wolfender, J.-L., Nuzillard, J.-M., van der Hooft, J. J. J., Renault, J.-H., and Bertrand, S. (2019). Accelerating metabolite identification in natural product research: toward an ideal combination of liquid chromatography-high-resolution tandem mass spectrometry and NMR profiling, in silico databases, and chemometrics. *Anal. Chem.* 91, 704–742. doi: 10.1021/acs.analchem.8b05112

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Rutz, Dounoue-Kubo, Ollivier, Bisson, Bagheri, Saesong, Ebrahimi, Ingkaninan, Wolfender and Allard. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Untargeted Metabolomics of *Nicotiana tabacum* Grown in United States and India Characterizes the Association of Plant Metabolomes With Natural Climate and Geography

Dong-Ming Ma¹, Saiprasad V. S. Gandra², Raman Manoharlal², Christophe La Hovary¹ and De-Yu Xie^{1*}

¹ Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC, United States, ² ITC Life Sciences and Technology Centre (LSTC), ITC Limited, Karnataka, Bengaluru, India

OPEN ACCESS

Edited by:

Kazuki Saito,
RIKEN Center for Sustainable
Resource Science (CSRS), Japan

Reviewed by:

Corey Broeckling,
Colorado State University,
United States
Akira Oikawa,
Yamagata University, Japan

*Correspondence:

De-Yu Xie
dxie@ncsu.edu

Specialty section:

This article was submitted to
Plant Metabolism
and Chemodiversity,
a section of the journal
Frontiers in Plant Science

Received: 05 May 2019

Accepted: 04 October 2019

Published: 30 October 2019

Citation:

Ma D-M, Gandra SVS, Manoharlal R,
La Hovary C and Xie D-Y (2019)
Untargeted Metabolomics of
Nicotiana tabacum Grown in United
States and India Characterizes the
Association of Plant Metabolomes
With Natural Climate and Geography.
Front. Plant Sci. 10:1370.
doi: 10.3389/fpls.2019.01370

Climate change and geography affect all the living organisms. To date, the effects of climate and geographical factors on plant metabolome largely remain open for worldwide and local investigations. In this study, we designed field experiments with tobacco (*Nicotiana tabacum*) in India versus USA and used untargeted metabolomics to understand the association of two weather factors and two different continental locations with respect to tobacco metabolism. Field research stations in Oxford, North Carolina, USA, and Rajahmundry, Andhra Pradesh India were selected to grow a commercial tobacco genotype (K326) for 2 years. Plant growth, field management, and leaf curing followed protocols standardized for tobacco cultivation. Gas chromatography–mass spectrometry based unbiased profiling annotated 171 non-polar and 225 polar metabolites from cured tobacco leaves. Principal component analysis (PCA) and hierarchical cluster analysis (HCA) showed that two growing years and two field locations played primary and secondary roles affecting metabolite profiles, respectively. PCA and Pearson analysis, which used nicotine, 11 other groups of metabolites, two locations, temperatures, and precipitation, revealed that in North Carolina, temperature changes were positively associated with the profiles of sesquiterpenes, diterpenes, and triterpenes, but negatively associated with the profiles of nicotine, organic acids of tricarboxylic acid, and sugars; in addition, precipitation was positively associated with the profiles of triterpenes. In India, temperature was positively associated with the profiles of benzenes and polycyclic aromatic hydrocarbons, but negatively associated with the profiles of amino acids and sugar. Further comparative analysis revealed that nicotine levels were affected by weather conditions, nevertheless, its trend in leaves was independent of two geographical locations and weather changes. All these findings suggested that climate and geographical variation significantly differentiated the tobacco metabolism.

Keywords: climate change, continent, geography, metabolomics, nicotine, tobacco, non-polar metabolite, polar metabolite

INTRODUCTION

The global warming trend has taken place over the last 100 years (Keller, 2007; Schnoor, 2007). With this continuous trend, the global mean temperature is predicted to increase by 1–4°C in the next 50–100 years (Jones, 2013; Kintisch, 2013). Accordingly, from last few decades global research efforts in different countries and regions have been undertaken to understand the effects of both global warming and regional climate changes on natural systems and human health. More endeavors are continuously being explored globally. The main research topics that have been studied and are being studied are focusing on the global scale impacts of climate change. These include effects of climate change on ecosystem and biodiversity, sea level rising, ocean acidification, water resources and desertification, agriculture and food security, and human health (Gosling et al., 2011). To date, all completed global research has led to numerous accomplishments which provide fundamental evidence that climate change significantly affects the global and regional ecosystems (Polley et al., 2013; Schmitz, 2013; Kizildenziz et al., 2015; Park et al., 2015; Hauser et al., 2016; Taylor and Kumar, 2016), agriculture, such as biotic and abiotic stresses (Adams et al., 1998; Dwivedi et al., 2013; Chauhan et al., 2014; Abberton et al., 2016; Kollah et al., 2016), and human health, such as the increase of infectious diseases and other public health problems (Patz et al., 1996; Rom and Pinkerton, 2014; Clayton et al., 2015; Kjellstrom et al., 2016). All of those data are fundamentally informative for different countries and global organizations to develop strategies and policies to slow down the global warming. To appropriately understand the effects of climate change on ecosystems, agriculture, and food safety, more endeavors are necessary to complete basic research at the molecular, biochemical, chemical, and physiological levels. However, this type of research is limited. For example, few studies have been reported to understand the effects of the combination of climate change and geography on natural product composition associated with crop and food safety and quality. Furthermore, there are particularly fewer reports regarding the comparative investigations of plant or crop metabolism associated with climate change across the continents. One of the main challenge is to conduct experiments using model plants or crops with the same genetic background that can be grown in the field in different continents. Accordingly, development of such a system is necessary to comprehensively enhance our understanding of the biological consequences of climate change.

Tobacco (*Nicotiana tabacum*) is an appropriate plant to study the effects of climate change on plant metabolism. As well characterized, metabolism in tobacco is complicated by biosynthetic networks of diverse plant metabolites such as alkaloids, terpenoids, phenolics, polyketides, alkanes, alkynes, and other metabolites, among which tobacco alkaloids, such as nicotine, cause public health problems (Sierro et al., 2014). Furthermore, more than 4,200 metabolites have been identified from tobacco plants and more than 8,000 metabolites in tobacco smoke (Rodgman and Perfetti, 2009). Currently, approximately 125 countries in Asia, Africa, North America, South America,

Europe, and Australia grow tobacco plants on more than 4.3 million hectares (Eriksen et al., 2015). One primary variety grown in all these regions is the flue-cured tobacco cultivar K326, and its genome sequencing was completed a few years ago (Sierro et al., 2014). All these advantages allow K326 to be an appropriate crop plant to study the effects of climate change and geographic region or a combination of both on plant metabolism.

In this study, we used K326 tobacco to conduct a 2-year pilot study in order to understand the effects of climate change and two continental growth locations on plant metabolism. Based on our previous metabolic profiling of a small number of tobacco samples collected from tobacco companies in India, North Carolina and Brazil (Ma et al., 2013), we hypothesized that weather and growing sites in different continents differentially altered the tobacco metabolomes associated with abiotic and biotic stresses. To test this hypothesis, we selected two field locations, the Oxford tobacco research station in North Carolina, USA, and the ILTD research station (department) of ITC limited at Rajahmundry, in Andhra Pradesh, India. These two locations have been used for long-term tobacco field research. We comparatively grew K326 tobacco plants in these two locations for two consecutive years. To understand the effects of climate on plant metabolism, we followed the standard tobacco industry protocols to manage field growth, collect, and cure leaf samples. We recorded daily temperature and precipitation conditions at these two locations. A large number of samples were collected for metabolic profiling of untargeted non-polar and polar metabolites using gas chromatography–mass spectrometry (GC-MS). A metabolite analysis pipeline was developed with Mass Professional Profile (Agilent Technologies, Inc). Metabolite profiles, main weather factors, and geographic locations were integrated for association analysis. The resulting data indicated an association of metabolic variations with climatic and geographical factors.

MATERIALS AND METHODS

Field Study Sites and Selection of Tobacco Variety

Two different continental field sites were selected for plant growth Viz., the Oxford research station in Oxford, North Carolina (latitude: 36, longitude: –78), USA and the ILTD research station of ITC limited in Rajahmundry, Andhra Pradesh, India (latitude: 16, longitude: 80). At each research station, three field plots were designed to grow plants, each plot had an area of 4047 m² (one acre) and constituted one field replicate (**Supplementary Figure 1**). The variety of *N. tabacum* cultivar used was K326, which is one of the primary flue-cured tobacco crops grown globally. The rationale was that this tobacco cultivar produces a large number of plant natural products and its field growth protocol has been industrially standardized for farmers in different continents. In addition, its secondary metabolism has been well documented in the literature. Therefore, it is an appropriate crop plant to investigate the effects of climate and geography on plant metabolism.

Experimental Design, Planting, and Sampling in North Carolina

Field growth of plants (**Supplementary Figure 1**) followed protocols documented for K326 in these two regions. The growth guide for K326 is published for farmers in North Carolina every year (<https://content.ces.ncsu.edu/catalog/collection/23069/flue-cured-tobacco-information>). We used this commercial growth guide to perform our experiments. Seeds were planted in soil contained in floating nursery beds in the greenhouse (**Supplementary Figure 2A**) at the end of February. This is the tobacco farming technique for commercial purpose. Seedlings were not removed out of the greenhouse until the second week of the May of 2011 (year 1) and 2012 (year 2), when plants were 72 days old. All seedlings grew strongly and uniformly (**Supplementary Figure 2B**). In North Carolina, prior to transplant, three plots were supplemented with 560 kg ha⁻¹ 8-8-24 base fertilizer and treated with clomazone and carfentrazone-ethyl/sulfentrazone for weed control. Each plot was planted with 800 plants (**Supplementary Figures 2C, D**) in the second week of May 2011 and 2012 (planting date was dependent upon weather). Thus, 2,400 plants in total were grown in each farming season. Plants were later side dressed with 168 kg ha⁻¹ of 15.5-0-0 fertilizer. Field growth management, such as irrigation (two to three times with irrigation reel and water from a lake) and the use of fungicides and herbicides, followed protocols documented in the research station. Plant growth was recorded and photographed each month (**Supplementary Figure 1**). When plants started to develop flower buds in early July, we selected 20 plants in each plot and then grouped 10 plants as one biological replicate. Consequently, each plot had two biological replicates. In total, 60 plants were selected from three plots and formed six biological replicates. All these plants almost grew the same height, flowered on the same date, and had 24 leaves. After this step, all plants were topped to keep 20 leaves from the bottom to the top and tagged group numbers from I to VII as shown in **Supplementary Figure 3**. After topping, plants were sprayed with maleic hydrazide to control suckers (axillary buds). After 3 weeks of topping, we started to harvest the leaf samples from group I at the bottom to group VII at the top (**Supplementary Figure 3**) on Aug. 3, Aug. 24, Sept. 14, and Oct. 4 (**Supplementary Table 1**). In total, we harvested seven groups of leaf samples, groups I–VII (**Supplementary Table 1** and **Supplementary Figure 3**). All samples were harvested in the morning. After the leaves were excised from the stems, were immediately separated into two halves, labeled as half 1 and half 2 (**Supplementary Figure 4**). The second half section included the main middle vein and was immediately cured in barn for metabolomics experiments. The first half section without the main middle vein was immediately frozen in liquid nitrogen, transported to laboratory by car, and then stored in –80°C freezers for other experiments. After 1 month of curing, 42 cured biological samples were obtained for metabolomics analysis in each year. In addition, 10 additional plants in each plot were selected and pooled as one biological sample for leaf harvest. However, leaves from these plants were not cut into two halves. Accordingly, seven uncut groups (I–VII, **Supplementary**

Figure 3) of leaves from 30 plants were harvested and immediately cured in the same barn. In total, 21 uncut biological samples were obtained (**Supplementary Table 2**). These samples were used as control to evaluate if cutting leaves into two halves could affect metabolite profiles. In summary, 84 cut leaf (biological) samples were collected from seven leaf positions (**Supplementary Figure 3**, **Supplementary Table 1**) and cured for metabolic profiling. To simplify the sample name description for data analysis, “U” was used to represent USA, and “F” and “S” were used to represent the first and second year of growth. Based on positions defined for sampling (**Supplementary Figure 3**), we used U-F1H through U-F7H to describe USA’s samples in the first year from positions 1 to 7 harvest. We also used U-S1H through U-S7H to describe USA’s samples in the second year from positions 1 to 7 Harvest. Each group of samples was composed of six biological replicates (**Supplementary Tables 1** and **3**). In addition, 42 entire leaf (un-cut, biological) samples as control were collected from seven positions (**Supplementary Table 2**) and cured for control metabolic profiling. Cured samples were ground into fine powder. For each biological sample, three technical replicates were prepared for non-polar and polar metabolite extraction described below.

Experimental Design, Planting, and Sampling in India

In India, field design, experimental design, and field management practices were the same as described above. However, field growth of plants and sampling times (**Supplementary Table 3**) were different from those in North Carolina because of tropical weather in India. In addition, plants were irrigated 6–8 times with ground water using a furrow irrigation method.

Moreover, the protocol developed for farmers requires to keep 22 leaves in each plant (**Supplementary Figure 5**). Therefore, in our experiments, after topping, each plant had 22 leaves, which were grouped into eight groups (**Supplementary Figure 5**). The leaf sampling approach was the same as described above for North Carolina’s samples (**Supplementary Figure 4**). In total, 48 biological leaf samples with a main vein (midrib) (**Supplementary Table 4**) were collected and cured for metabolomics in each year. Samples frozen in liquid nitrogen were transported to ITC laboratory by air flight and then stored in –80°C freezers. In addition, 24 biological samples of entire leaves without cutting were harvested and then cured as control for metabolomics in each year. Daily weather conditions during two growing seasons were obtained from local weather station. In summary, in 2 years of growth, 96 cut leaf (biological) samples were collected from eight positions (**Supplementary Table 3**) and cured for metabolomics. To simplify the sample’s name description used for data analysis, we used “I” to represent India (**Supplementary Tables 3**). We used I-F1H through I-F8H to describe India’s samples in the first year from positions 1 to 8 harvest. We also used I-S1H through I-S8H to describe India’s samples in the second year from positions 1 to 8 harvest. Each group of samples included six biological replicates (**Supplementary Tables 3**). In addition, 48 entire leaf (un-cut, biological) samples were collected from eight positions (**Supplementary Table 4**) and

cured as control. Cured samples were ground into fine powder. Powdered samples were then shipped to North Carolina State University for metabolic profiling. For each biological sample, three technical replicates were prepared for non-polar and polar metabolite extraction described below.

Extraction of Non-Polar Metabolites

Extraction of non-polar metabolites was performed as reported previously (Ma et al., 2013). In brief, 100 mg of powdered sample was suspended in 1.5 ml 100% hexane in a 2 ml Eppendorf tube. The extraction tube was vortexed vigorously for 1 min, followed by 30 min sonication, and then placed in a 56°C water bath for additional 50 min. The extraction tube was centrifuged at 11,000 rpm for 10 min. The resulting clean hexane supernatant was pipetted into a new clean tube. A 200 µl hexane extract was pipetted into a 400 µl insert contained in a 2 ml glass vial (Agilent) for gas chromatograph-mass spectrometry analysis described below.

Extraction of Polar Metabolites and Derivatization

One hundred milligram of powdered sample was suspended in 1.5 ml extraction solvent (methanol: chloroform: water, 5:2:2) in a 2.0 ml Eppendorf tube and then vigorously vortexed for one min. The extraction tube was sonicated for 30 min in water bath and placed in a 56°C water bath for 50 min. The extraction tube was centrifuged at 11,000 rpm for 10 min. One milligram of the supernatant was transferred to a new clean 1.5 ml tube, to which 300 µl chloroform and 600 µl double deionized H₂O were added. The tube was vigorously vortexed for 1 min to mix solvents thoroughly, followed by centrifugation at 4,000 rpm for 5 min to result in the upper water-methanol (polar) phase and the lower chloroform (non-polar) phase. One hundred microliter water-methanol phase was transferred to a new 1.5 ml Eppendorf tube and then was evaporated in a rotary vacuum at room temperature. The remaining residues at the bottom of tubes contained polar metabolites, which were used for derivatization as reported previously (Ma et al., 2013). In brief, the remaining residue was oximated using 40 µl methoxylamine hydrochloride (20 mg/ml) in anhydrous pyridine at 37°C for 2 h, and then were silylated at 37°C for 30 min using 70 µl of N-methyl-N-(trimethylsilyl)-trifluoro acetamide (MSTFA). The liquid derivatized samples were pipetted into 400 µl glass vials for GC-MS analysis as reported (Duan et al., 2012), which is described below.

Gas Chromatograph-Mass Spectrometry Analysis

An Agilent 6890 Gas Chromatograph coupled with 5975 MSD (Agilent Technologies, USA) was used to profile metabolites extracted from 180 cut leaf samples (84 from North Carolina and 96 from India) and 90 (42 from North Carolina and 48 from India) uncut entire leaf samples (Supplementary Tables 1–4). Prior to injection of samples, we carefully performed instrumental quality assurance and quality control (QA/QC), including but not limiting to the use of new consumables, cleaning of systems, tuning, blank solvent control, and others. Three technical replicates were prepared

for each biological sample, of which two technical replicates were analyzed. The third technical replicate was assayed for some of samples when it was necessary. As a result, 12 or 18 repetitive assays for each group leaf sample (such as U-F1H from group I, Supplementary Figures 3 and 4) were carried out to profile non-polar and polar metabolites. A RTX-5 capillary column (30 m×0.25 mm×0.25 µm) was used to separate metabolites. The inlet was operated using a splitless mode. One microliter sample was injected to profile metabolites. The injection temperature was set at 250°C. The temperature of column oven was initially set at 60°C for 2 min, then ramped to 320°C at a constant rate of 8°C/min, and held at 320°C for 2 min. Pure helium was used as the carrier gas with a flow rate of 1 ml/min. A positive electron impact ion source (70 EV) was used to ionize compounds and mass fragments were scanned in the range of 40–800 (m/z) starting with 4 min of retention time. We recorded total ion chromatographs for all samples. In addition, a mixture of even-numbered chain n-alkenes (C10–C40) purchased from Restek (Florida, USA, catalog no. 31266) were used to estimate the retention index values, which were used to deconvolute metabolite peaks as reported previously (Almaarri et al., 2010).

Data Processing and Statistical Analysis

GC-MS ChemStation data format for metabolites detected from each assay was translated into the MassHunter data format using Agilent MassHunter GC/MS Translator Software (version B.05.02). Then, untargeted analysis was carried out using the Mass Hunter Qualitative Analysis software (Version B.06.00) for metabolite deconvolution, compound search, and identification in NIST11 library. When the MS profile of an untargeted metabolite showed no less than 80% identity to that of a standard metabolite in the library, it was annotated to the standard. Then, the MassHunter data format of all annotated metabolites was exported as Compound Exchange Format (.cef). All “cef” files were imported to the Mass Profiler Professional (MPP, version B.12.5, Agilent) for statistical analysis.

The resulting “cef” files for all samples were imported to MPP for principal component analysis (PCA) and hierarchical clustering analysis (HCA) according to our previous report (Ma et al., 2015). For PCA, we used Eigen-vector based scaling method to perform alignment and normalization (using log₂) and then visualize metabolite profiles in all samples. The abundance of metabolites was evaluated using counts. In MPP, the minimum abundance count and retention time tolerance (min) values across all sample sets were 5,000 and 0.05, respectively. These threshold values were used as a filter to determine the presence or absence of metabolites in samples. In addition, HCA is a statistical method to group samples that are unsupervised in different clusters or branches of the hierarchical tree. The MPP software processing tool is effective to align all the data and then normalize them using log₂ for each metabolite. The MPP identified the median value across all the biological samples (180 cut leaf samples) and then baselined each metabolite account value to the median value to establish the hierarchical clustering (conditional tree) and heat maps. The resulting tree showed the relationships between different groups of samples from both North Carolina and India.

Pearson Analysis and Principal Component Analysis for Metabolites, Temperature and Precipitation in North Carolina and India Alone

Pearson analysis was performed using the SPSS software (IBM SPSS Statistics 22) to understand the potential relevance of one or two climate factors and metabolite profiles of leaves from in both India and North Carolina. Weather stations in Oxford County in North Carolina, USA, and Rajahmundry in Andhra Pradesh, India recorded local field air temperature and daily precipitation. We obtained these data from weather stations and then calculated the average values for every week of plant growth. To perform Pearson analyses, variables that were used included average temperature and precipitation values in each harvest week, nicotine (a main tobacco alkaloid in leaves) peak values, and peak values of 11 groups of metabolites (including six groups of non-polar and five groups of polar metabolites, **Supplementary Table 9**). The six non-polar metabolite groups included monoterpenes, sesquiterpenes, diterpenes, triterpenes, benzene, and polycyclic aromatic hydrocarbons (PAH). The five polar metabolite groups included sugars, amino acids, organic acid, nitrogen-containing secondary metabolites (nicotine-related compounds) and polyphenol. The level of each group resulted from the sum of peak values of all individual metabolites. For example, the monoterpene level was summed from all individual molecule peak values. The resulting average temperature values per week, average precipitation values per week, and total peak values of each group of metabolites were input to the SPSS software for Pearson analysis.

Eigen-vector based PCA was carried out with the JUMpro 12 software (North Carolina State University, Raleigh, USA) to understand the potential relevance between 12 groups of metabolites (nicotine and 11 groups), average daily temperature, and average daily precipitation at locations in both India and North Carolina. The PCA type used during the computation was the Pearson's correlation matrix. This analysis corresponds to the classical correlation coefficient. The resulting plots were characterized by arrows that represented observations and variables simultaneously. When two variables were distant from the plot center, there were three results. The first is that two are close to each other, indicating that they were significantly positively correlated (r close to 1). The second is that two are orthogonal each other, indicating that they are not correlated (r close to 0). The third is that two are opposite to each other, indicating that they are significantly negatively correlated (r close to -1).

RESULTS

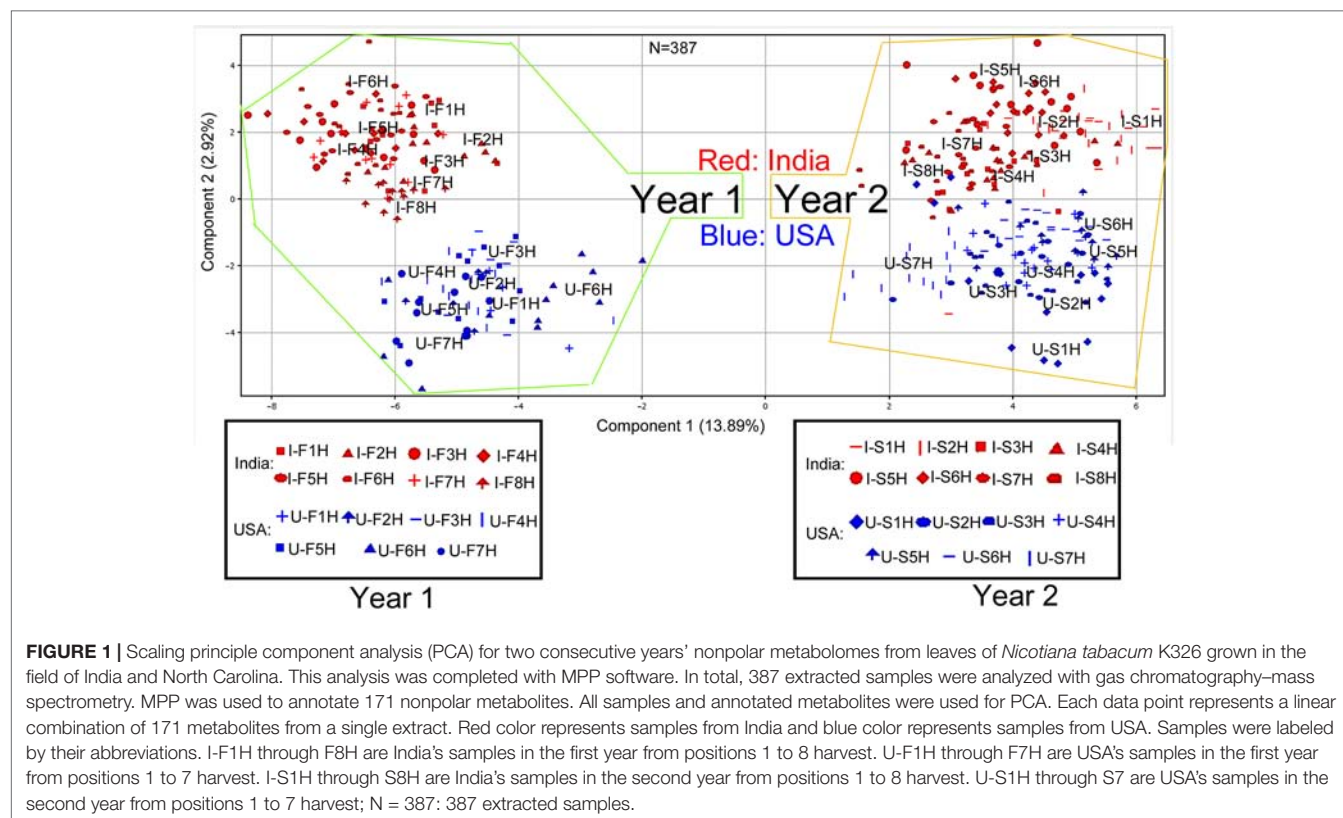
Non-Polar Metabolite Profile Complexity in Leaves From North Carolina and India

We used GC-MS to complete 387 assays for 180 biological samples. These included 84 and 96 cut leaf (biological) samples collected from 2 years' field growth studies in North Carolina and India, respectively. The 84 biological samples from North Carolina were composed of seven groups of leaves and the 96 biological samples

from India were composed of eight groups of leaves. Each group of samples was composed of six biological replicates. Two technical replicates for each biological sample were analyzed using GC-MS. In addition, 27 biological samples were randomly selected from 180 samples for a third technical replicate analysis in order to confirm experimental reproducibility of GC-MS. As a result, GC-MS analysis detected more than 700 different metabolite peaks from 387 extractions (**Supplementary Table 5**). When a peak was detected from at least 20 samples, it was considered being a positive one. The peak was then annotated as a compound with at least 80% identity to a standard in the library. Accordingly, 171 non-polar metabolites (**Supplementary Table 6**) were annotated from 387 assays. Based on skeleton structure features, these metabolites were classified into nine groups, including: linear hydrocarbons, cyclic hydrocarbons, alcohol, aldehyde, ketones, acids, terpene, polyphenol, and nitrogen-containing natural products.

All 171 metabolites annotated and their ion chromatographic peak values, growing years, and locations were used as entries for PCA in MPP. The resulting two-dimensional ordination plot showed that the first principal component (PC1) and the second principal component (PC2) accounted for 13.89 and 2.92% of the total variance of metabolites across 387 extracted samples from two different years across two growth locations (**Figure 1**). Based on this plot, the ordination orders of two growing years and locations were separated by variables consisting of 171 metabolite and 387 extracted samples. Regardless of the locations, the growing years were ordinally separated in the PC1 axis, indicating differential effects of growing years on the profiles of 171 metabolites. In each year, North Carolina (USA) and India were ordinally separated in the PC2 axis, indicating the difference of metabolism of these 171 metabolites in two growing locations.

The abovementioned 171 metabolites and their ion chromatographic peak values, growing years, and locations were also used for HCA. As a result, HCA created two dendrograms (I and II), two heat maps (1 and 2), and two color bars (I and II) (**Figure 2**). These graphic features allowed visualizing the clustering distance between paired samples in two growth places, e.g., I-F1H vs. U-F1H, and in 2 years, e.g., I-F1H vs. I-S1H. In addition, these graphs revealed clustering distance between sample groups from the bottom to the top of plants, e.g., I-F1H vs. I-F2H. These data showed effects of two growing years and two locations on 171 metabolite profiles from 387 extracted samples (30 groups of samples). The color bar I, heat map I, and dendrogram I characterized that based on metabolite profiles, samples in each of the first and second years alone was clustered, respectively, regardless of the field location in India and North Carolina. These data were consistent with the results of PCA described above. These data also indicate that the growing years affect the non-polar metabolite complexity in plants. The heat map I and dendrogram I also showed that under each year, samples from India and samples from North Carolina were grouped together, respectively. These results indicate that in the same year, the non-polar metabolite complexity in samples depend upon each growth location. The color bar II, heat map II, and dendrogram II categorize metabolites from high to low abundance in all samples regardless of growing locations and years. The color bar II from deep red to deep blue characterizes



the peak abundance of the 171 metabolites shown in heat map II, each of which is coded by different colors in the direction of the black arrow (**Figure 2**). Based on color codes, we categorized the 171 metabolites into three groups (G-i, ii, and iii). The G-i is highly abundant in all 387 extracted samples regardless of growing years and locations. Example metabolites of this group includes squalene, vitamin E, stigmasterol, some diterpenes [such as 4,8,13-cyclotetradecatriene-1,3-diol, 4,8,13-cyclotetradecatriene-1,3-diol, 1,5,9-trimethyl-12-(1-methylethyl)-(C₂₂H₃₄O₂), and phytol acetate]. The G-ii coded by pink to yellow-brownish color in the clustering are middle abundant metabolites, including nicotine and 1,3-bis(1-formylethyl) benzene. The G-iii coded by yellowish to bluish color are low abundant metabolites. Examples include monoterpenes (such as limonene, 1,7,7-trimethyl-bicyclo[2.2.1]hept-2-ene, and 1,2-dimethyl-3-pentyl-cyclopropane), sesquiterpenes [such as, solavetivone, 9-(3,3-dimethyloxiran-2-yl)-2,7-dimethylnona-2,6-dien-1-ol, 1S,4R,7R,11R-1,3,4,7-tetramethyltricyclo(5.3.1.0{4,11})undec-2-en-8-one], diterpenes (such as thunbergol, andrographolide, and others), triterpenes, 1-butylheptyl-benzene, and decahydro-2,3-dimethyl-naphthalene.

Polar Metabolite Profile Complexity in Leaves From North Carolina and India

We used GC-MS to complete total 366 assays for 180 biological samples to analyze polar metabolites. Two technical replicates were analyzed for each biological replicate. In addition, six

biological samples were randomly selected from 180 samples for a third technical replicate to examine technical reproducibility. More than 500 metabolite peaks were detected from 366 assays (**Supplementary Table 7**). Based on 80% identity criteria, 225 peaks (**Supplementary Table 8**) were annotated to metabolites using the MPP software. These compounds were categorized into seven different groups, including sugar, amino acid, organic acid, alcohol and ketone, N-containing natural products, hydrocarbons, and phenylpropanoids (**Supplementary Table 8**).

All 225 metabolites annotated and their ion chromatographic peak values, 366 extracted samples, two growing years, and two locations were used as variables for PCA. The resulting ordination plot showed that the percentages of PC1 and PC2 were 18.13 and 7.96% across all variables (**Figure 3**). Regardless of growth locations, the resulting plot showed that the polar metabolome of the 1st year's samples was ordinally separated from that of the 2nd year's samples in the PC1 axis (**Figure 3**), indicating that growing years obviously altered the plant metabolism. In addition, the resulting plot revealed that samples from India and USA were ordinally separated in the PC2 axis (**Figure 3**), indicating different metabolism of these compounds in two growing locations. Based on this plot, the profiles of these 225 metabolites between North Carolina and India were more similar in the first year's leaf samples (U-F1H to U-F7H and I-F1H to I-F8H) than in the 2nd year's leaf samples (U-S1H to U-S7H and I-S1H to I-S8H). It was interesting that in the second year, dynamic metabolite profile distributions were observed between North Carolina and India samples. In the 2nd year, metabolite profiles in seven groups

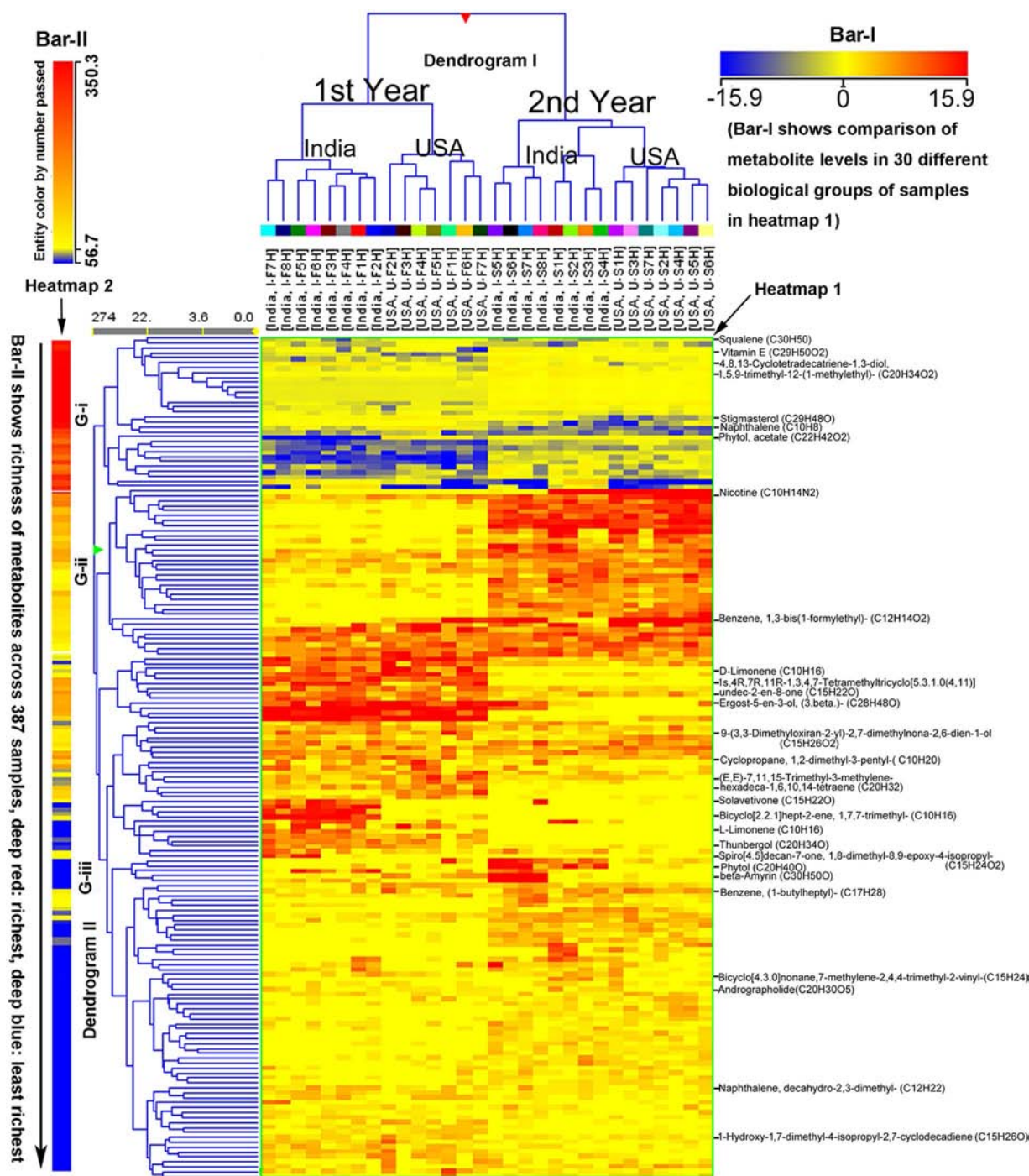


FIGURE 2 | Hierarchical cluster analysis (HCA) for two consecutive years' nonpolar metabolomes from leaves of *Nicotiana tabacum* K326 grown in the field of India and North Carolina. A heat map resulted from HCA using log₂-fold changes of 171 metabolites during two growing seasons in India and USA. Fold changes and color (bar-I) resulted from normalization that was carried out using each log₂ value of each metabolite peak value in each sample to compare its median peak value across all 387 extracted samples. Two color bars are used in the heat map. The horizontal bar-I fold color bar (15.9 to -15.9) from deep red to deep blue shows high to low levels of each metabolite in different groups of biological samples. The bar-II entity color by number passed (350.3–56.7) and the long arrow from the top to the bottom featured using deep red to deep blue color indicate the frequency of metabolites detected from the highest (deep red) to the lowest (deep blue) richness in all 387 extracted samples. For example, squalene listed in the top of the heat map was the most abundant metabolite detected in all samples, while naphthalene, decahydro-2,3-dimethyl- listed in the relative bottom with deep blue color was one of the least abundant metabolite detected in all samples. Samples were labeled by their abbreviations. I-F1H through F8H are India's samples in the first year from positions 1 to 8 harvest. U-F1H through F7H are USA's samples in the first year from positions 1 to 7 harvest. I-S1H through S8H are India's samples in the second year from positions 1 to 8 harvest. U-S1H through S7 are USA's samples in the second year from positions 1 to 7 harvest; N = 387: 387 extracted samples.

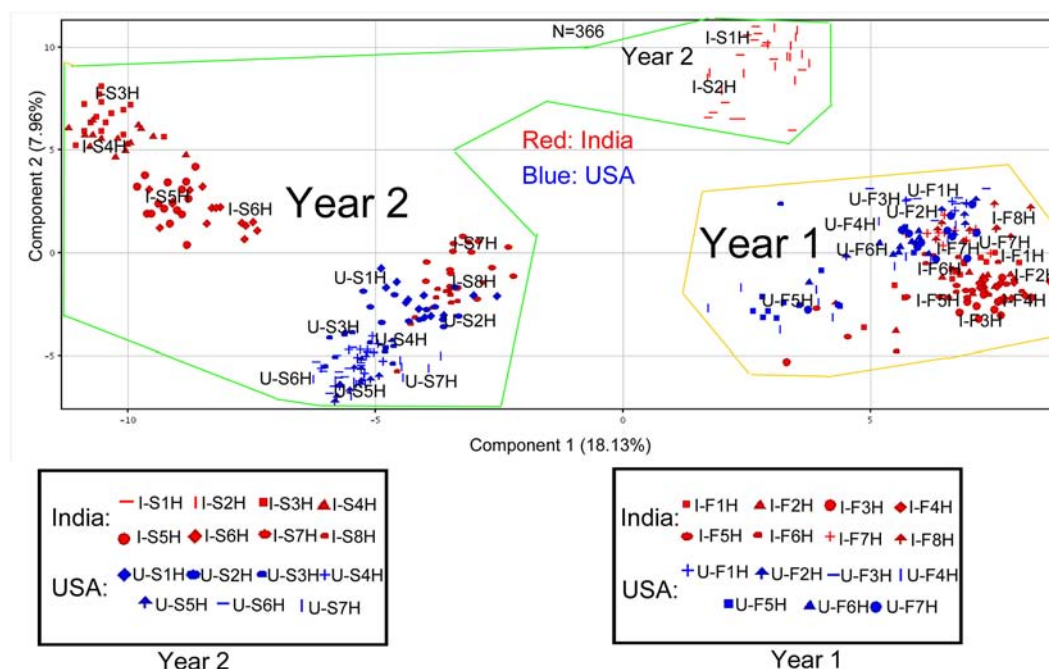


FIGURE 3 | Scaling principle component analysis (PCA) for two consecutive years' polar metabolomes from leaves of *Nicotiana tabacum* K326 grown in the field of India and North Carolina. This analysis was completed with MPP software. In total, 366 extracted samples were analyzed by gas chromatography–mass spectrometry. MPP was performed to annotate 225 peaks to polar metabolites. All samples and annotated metabolites were used for PCA. Each data point represents a linear combination of 225 metabolites from a single extract. Samples were labeled by their abbreviations. I-F1H through F8H are India's samples in the first year from positions 1 to 8 harvest. U-F1H through F7H are USA's samples in the first year from positions 1 to 7 harvest. I-S1H through S8H are India's samples in the second year from positions 1 to 8 harvest. U-S1H through S7 are USA's samples in the second year from positions 1 to 7 harvest. N = 366: 366 extracted samples.

of samples (U-S1H to U-S7H) from North Carolina were closely grouped together, while the profiles of these metabolites in eight groups of samples from India (I-S1H to I-S8H) were ordinally separated into three sections in the PC1 axis (Figure 3).

The abovementioned 225 metabolites and their peak values, 366 extracted samples, two growing years, and two locations, were also used as variables for HCA. As described above, HCA created two heat maps (I and II), two dendrograms (I and II), and two color bars (I and II) to visualize metabolite profiles in 366 extracted samples (Figure 4). Bar I, heat map I, and dendrogram I characterized metabolite profiles in 366 extracted samples (30 groups of samples) from 2 years. The heat map I and dendrogram I show that 15 groups of the first year's samples (eight from both India and seven from North Carolina) are clustered together, while 13 groups of the second year's samples (seven from both North Carolina and six from India) are clustered together. Two groups of samples in the second year, I-S1H and I-S2H from India, are clustered as an out-group, which is clustered together with the first year's samples. These results were in agreement with those obtained from PCA, in which these two groups of samples were relatively ordinally distant from 13 sample groups of the second year in the first PC1 axis (Figure 3). These results indicate that metabolic accumulation of plants are associated with the two growing years. Based on the tree in dendrogram I, in each year, samples from India and North Carolina were separately clustered together. These results indicate that in the same year, growth locations control metabolic complexities. Bar II, heat map II,

and dendrogram II characterized peak abundance profiles of the 225 metabolites in 366 extracted samples regardless of years and locations. Based on color codes, we categorize the 225 metabolites into three groups, G-i: high abundance, G-ii: middle abundance, and G-iii: low abundance, which are colored by deep red to orange, orange to yellow, and blue, respectively (Figure 4). Examples of G-i include glutamic acid, malic acid, fumaric acid, succinic acid, aspartic acid, serine, phenylalanine, threonine, and others. Examples of G-ii include glucose, arabinose, fructose, and others. Examples of G-iii include xylose, 1'-demethyl nicotine, and other compounds (Figure 4).

Association of Two Climatic Factors and 12 Groups of Metabolites in North Carolina

Both air temperature and precipitation of two growing seasons in North Carolina were recorded daily. Average values for each leaf harvest week were calculated. The dynamic trend of average temperature values was similar at Oxford research station in 2 years. The values decreased from the first (Aug. 3) to the last (Oct. 4) harvest week (Figure 5A). The average temperature values were higher in the 1st (Aug. 3), 2nd (Aug. 14), 3rd (Aug. 24) and 5th (Sept. 14) weeks of harvest in the 1st year than in the 2nd year (Figure 5A). The dynamics of average weekly precipitation values were different in the 2 years of growth (Figure 5B). During the seven continuous weeks of harvest, the average precipitation values in the 3rd (Aug. 24), 5th (Sept. 14) and 7th (Oct. 4) were relatively close, but were

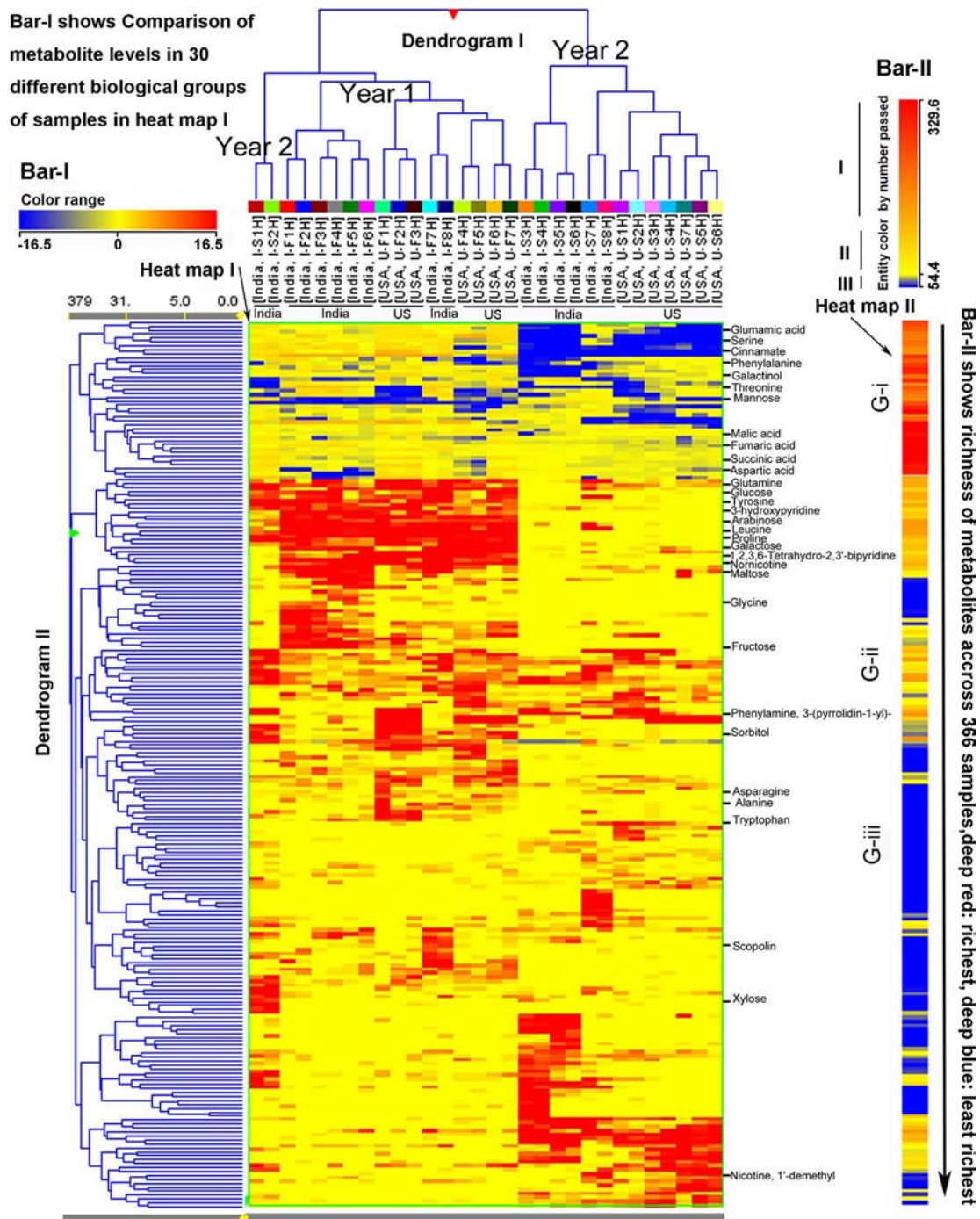


FIGURE 4 | Hierarchical cluster analysis (HCA) for two consecutive years' polar metabolomes from leaves of *Nicotiana tabacum* K36 grown in the field of India and North Carolina. A heat map resulted from HCA using log₂-fold changes of 225 metabolites during two growing seasons in India and USA. Fold changes and color (bar-I) resulted from normalization that was carried out using each log₂ value of each metabolite peak value in each sample to compare its median peak value across all 366 extracted samples. Two color bars, bar-I and bar-II, are used in the heat map. The horizontal bar-I fold color bar (16.5 to -16.5) from deep red to deep blue shows high to low levels of each metabolite in different groups of biological samples. The bar-II entity color by number passed (329.6–54.4) and the long arrow from the top to the bottom featured using deep red to deep blue color indicate the frequency of metabolites detected from the highest (deep red) to the lowest (deep blue) richness in all 366 extracted samples. For example, glutamic acid listed in the top of the heat map was the most abundant metabolite detected in all samples, while xylose listed in the relative bottom with deep blue color was one of the least abundant metabolite detected in all samples. Samples were labeled by their abbreviations. I-F1H through F8H are India's samples in the first year from positions 1 to 8 harvest. U-F1H through F7H are USA's samples in the first year from positions 1 to 7 harvest. I-S1H through S8H are India's samples in the second year from positions 1 to 8 harvest. U-S1H through S7 are USA's samples in the second year from positions 1 to 7 harvest. N = 366: 366 extracted samples.

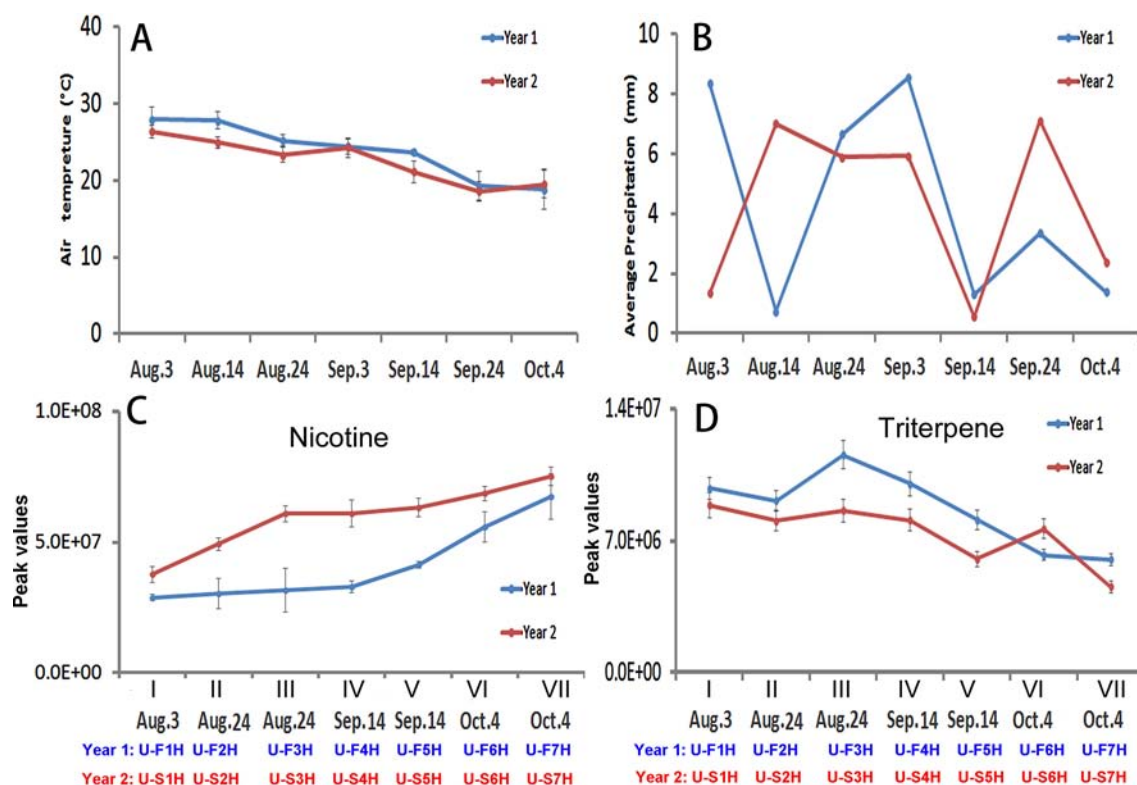


FIGURE 5 | Dynamics of two climate factors and trends of two groups of metabolites in two harvesting seasons in North Carolina. **(A–B)**, Dynamics of average temperature **(A)** and precipitation **(B)** values at seven harvesting dates. Each value resulted from averaging daily temperature or precipitation levels in that week of harvest. **(C–D)** Trends of total nicotine levels **(C)** triterpene levels **(D)** in seven groups of samples harvested on the four harvesting dates. I, II, III, IV, V, VI, and VII: seven groups of samples, in which II and III were collected on Aug. 24, IV and V were collected on Sept. 14, and VI and VII were collected on Oct. 4.

higher in the 1st and 4th, and lower in the 2nd and 6th week of harvest in the first year than in the second year.

The level trends of nicotine levels and 11 groups of metabolites (triterpenes, monoterpenes, diterpenes, sesquiterpenes, benzene, PAH, amino acid, organic acids of the tricarboxylic acid (TCA) cycle, sugars, nitrogen containing compounds, and polyphenols) in seven groups of leaves were characterized using their peak values (**Figures 5C, D; Supplementary Figure 6**). Except for nicotine, the levels of each group of metabolites were summed from individuals, such as the level of monoterpenes summed from all individuals. The resulting plots showed three types of trends in 2 years. The first trend was similar in 2 years but the levels of group metabolites at each harvest date were lower in the first year than in the second year. This type included nicotine and benzenes. The trends of nicotine in 2 years of growth were similar from the lowest level in the group I leaves to the highest in the group VII leaves (**Figure 5C**). In comparison, the levels of nicotine in each group of leaves at each harvest date was higher in the second year than in the first year. The trends of benzenes were also similar in 2 years. The levels of benzenes were lower in each group of leaves at each harvest date in the first year than in the second year (**Supplementary Figure 6D**). The second type was that the levels of groups of metabolites in all or most samples groups (I–VII) were higher in the first year than in the second year. These groups of metabolites included triterpenes (**Figure 5D**), monoterpenes, diterpenes, amino acids,

N-containing compounds (natural products), and polyphenols (**Supplementary Figures 6A, C, F, I, and J**). The accumulation patterns of triterpenes showed a trend from the highest in the group I leaves to the lowest in the group VII samples in 2 years (**Figure 5D**). Except for the group VI leaves, the levels of triterpenes in leaves were higher in the second year than in the first year. The levels of monoterpene, amino acids, and polyphenols in each group of samples at each harvest date were higher in the first year than in the second year. The third type trends for sesquiterpenes, PAH, organic acid, and sugar (**Supplementary Figures 6B, E, G, H**) were dynamic in the seven groups of samples between 2 years. Taken together, these results indicate that the metabolism of 12 groups of metabolites differentially respond to the growing years.

These 12 groups of metabolites, seven groups of samples, and 2 years were used as variables for PCA (with JUMpro software) to understand the potential correlation of growing years and metabolite profiles. The resulting two-dimensional ordination plot showed that the PC1 and the PC2 accounted for 41.9 and 23.8% of the total variance from 2 years in North Carolina (**Figure 6A**). In general, the seven group of samples were separated between the first and second years. The ordination plot also showed that in each year, the ordination order of seven group samples was relatively similar from I to VII. These data indicate that the positions (groups) of leaves and harvest times are associated with 12 groups of metabolite profiles.

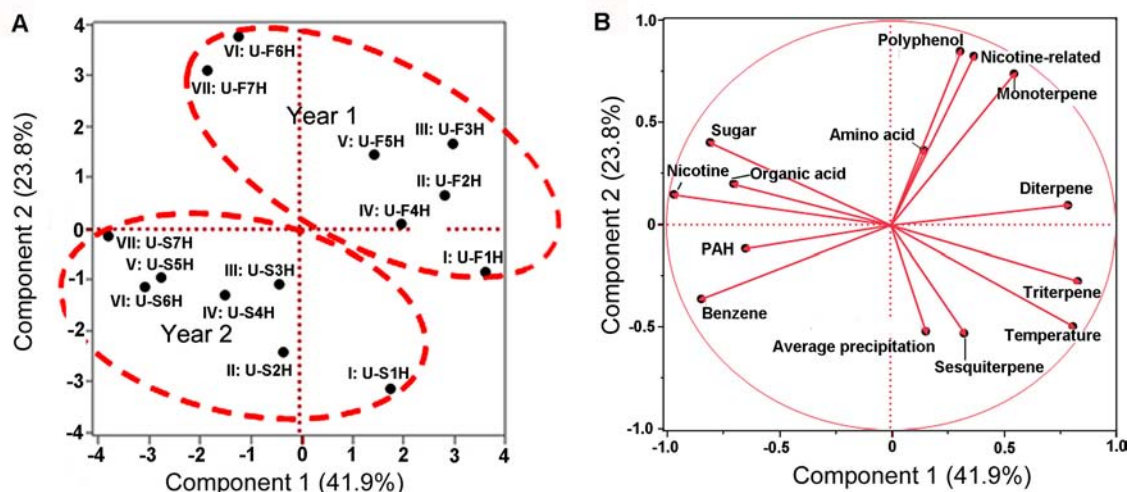


FIGURE 6 | Scaling principal component analysis (PCA) showing associations between metabolite profiles with growing years and two climatic factors in North Carolina. This analysis was completed with JUMpro software. **(A)**, A plot of PCA shows an association between two growing years and metabolite profiles in seven groups of samples. In general, group samples showed a similar tendency in two consecutive years. **(B)**, A plot of PCA shows associations between two climatic factors and 12 groups of metabolites (including nicotine) as well as correlation among 12 groups of metabolites themselves in 2 years of field growth.

These 12 groups of metabolites and two climatic factors (average temperature and precipitation values) were used as variables for PCA to understand ordination relevance. The resulting PCA plot showed an ordinate relevance between two climate factors and 12 groups of metabolites (Figure 6B). Temperature, precipitation, triterpene, and sesquiterpene fell in the same ordinate region, suggesting a close positive association. By contrast, nicotine, sugar, and organic acid fell in the opposite ordinate region against both temperature and precipitation, suggesting a negative ordinate relevance. Average precipitation also had a positive association with triterpene and sesquiterpene but had a negative relevance to nicotine, sugar, and organic acid. Other groups of metabolites were localized in two other ordinate regions, suggesting a potential partial relevance to these two weather factors.

Furthermore, a Pearson analysis was performed to evaluate the correlation between 12 groups of metabolites and these two climatic factors established by PCA. A bi-variate Pearson analysis revealed a linear relationship among variables (Table 1). The resulting data showed that temperature was positively associated with triterpene (P-value less than 0.01), sesquiterpene (P-value less than 0.05), and diterpene (P-value less than 0.05). By contrast, the resulting data showed that temperature was negatively associated with nicotine (P-value less than 0.01), sugar (P-value less than 0.01), and organic acid of TCA (P-value less than 0.05). The resulting data also showed a significantly positive association between average precipitation and triterpene (P-value less than 0.01).

Association of Two Climatic Factors and 12 Groups of Metabolites in India

As described above, daily temperature and precipitation values were recorded at the field in the two growing seasons in India. The average temperature values were calculated for each harvest

week. During the entire growing seasons, the average temperature values showed an increased trend from Feb. 9 through March 31 in both years (Figure 7A). In the week of Feb. 9, Feb. 23, and Mar. 31, average temperature values were very close in 2 years, while the temperature values in other weeks of March were higher in the second year than in the first year. Regarding daily precipitation, it was mostly dry during the harvest period in the two growing years (Figure 7B). From Feb. 9 to March 31, it rained only once in each year, on Feb. 20 in the first year and on March 13 in the second year (Figure 7B).

The trends of nicotine and 11 groups of metabolites were analyzed with their total peak values in eight groups (different positional) of leaves collected on eight dates from Feb. 9 through Mar. 31. The resulting plots showed three types of dynamic trends. The first trend was that the levels of four groups of metabolites in all groups (at all dates) or groups V–VIII (at the late half dates) were higher in the second year than in the first year. Nicotine, triterpenes, benzenes, and PAH followed this trend (Figures 7C, D and Supplementary Figures 7D and E). The nicotine levels in the eight groups of leaves showed an increased trend in two years (Figure 7C). In comparison, the levels of nicotine in each group of leaves were higher in the second year than in the first year. The second trend was that the total levels of four groups of metabolites in all groups (all dates) or most groups (most dates) samples were higher in the first year than in the second year. These groups of metabolites included monoterpenes, sesquiterpenes, diterpenes, and polyphenols (Supplementary Figures 7A–C, I). The third trend was that the levels of four groups of metabolites were dynamic in eight groups of samples in the 2 years. These includes, amino acids, organic acids, and sugars (Supplementary Figures 7F, G, H, and I). Although the levels of sugars were dynamic, their values increased from I-F1H and I-S1H through I-F8H and I-S8H (Supplementary Figure 7H).

TABLE 1 | Data of Pearson correlation analysis for two weather facts and 12 groups of metabolites in North Carolina.

		Temperature	Daily precipitation	Monoterpene	Sesquiterpene	Diterpene	Triterpene	Benzene	PAH	Nicotine	Amino acid	Organic acid	Sugar	NCSM	Polyphenol
Temperature	Pearson correlation	1	0.225	0.057	0.556*	0.582*	0.767**	−0.519	−0.465	−0.855**	−0.119	−0.655*	−0.855**	−0.092	−0.176
	Sig. (two-tailed)		0.439	0.847	0.039	0.029	0.001	0.057	0.094	0	0.685	0.011	0	0.754	0.547
Daily precipitation	Pearson correlation	0.225	1	−0.328	0.194	0.08	0.553*	0.115	−0.106	−0.216	−0.032	0.16	−0.356	−0.309	−0.206
	Sig. (two-tailed)	0.439		0.252	0.507	0.786	0.040	0.695	0.719	0.458	0.914	0.584	0.212	0.283	0.480
Monoterpene	Pearson correlation	0.057	−0.328	1	−0.338	0.478	0.226	−0.747**	−0.33	−0.462	0.13	−0.376	−0.169	0.710**	0.819**
	Sig. (two-tailed)	0.847	0.252		0.237	0.084	0.436	0.002	0.249	0.096	0.657	0.186	0.565	0.004	0
Sesquiterpene	Pearson correlation	0.556*	0.194	−0.338	1	0.017	0.25	−0.009	−0.443	−0.357	0.099	−0.206	−0.511	−0.105	−0.285
	Sig. (two-tailed)	0.039	0.507	0.237		0.955	0.39	0.976	0.112	0.211	0.737	0.479	0.062	0.72	0.324
Diterpene	Pearson correlation	0.582*	0.08	0.478	0.017	1	0.722**	−0.604*	−0.222	−0.775**	0.267	−0.591*	−0.532	0.361	0.255
	Sig. (two-tailed)	0.029	0.786	0.084	0.955		0.004	0.022	0.446	0.001	0.356	0.026	0.05	0.205	0.379
Triterpene	Pearson correlation	0.767**	0.553*	0.226	0.25	0.722**	1	−0.571*	−0.388	−0.835**	0.104	−0.491	−0.752**	0.083	0.106
	Sig. (two-tailed)	0.001	0.04	0.436	0.39	0.004		0.033	0.171	0	0.722	0.074	0.002	0.777	0.717
Benzene	Pearson correlation	−0.519	0.115	−0.747**	−0.009	−0.604*	−0.571*	1	0.718**	0.734**	−0.075	0.618*	0.452	−0.515	−0.543*
	Sig. (two-tailed)	0.057	0.695	0.002	0.976	0.022	0.033		0.004	0.003	0.799	0.019	0.105	0.06	0.045
PAH	Pearson correlation	−0.465	−0.106	−0.33	−0.443	−0.222	−0.388	0.718**	1	0.503	−0.237	0.274	0.45	−0.389	−0.314
	Sig. (two-tailed)	0.094	0.719	0.249	0.112	0.446	0.171	0.004		0.066	0.415	0.343	0.106	0.17	0.274
Nicotine	Pearson correlation	−0.855**	−0.216	−0.462	−0.357	−0.775**	−0.835**	0.734**	0.503	1	−0.01	0.747**	0.852**	−0.239	−0.179
	Sig. (two-tailed)	0	0.458	0.096	0.211	0.001	0	0.003	0.066		0.974	0.002	0	0.411	0.54
Amino acid	Pearson correlation	−0.119	−0.032	0.13	0.099	0.267	0.104	−0.075	−0.237	−0.01	1	0.262	−0.118	0.567*	0.242
	Sig. (two-tailed)	0.685	0.914	0.657	0.737	0.356	0.722	0.799	0.415	0.974		0.366	0.689	0.035	0.404
Organic acid	Pearson correlation	−0.655*	0.16	−0.376	−0.206	−0.591*	−0.491	0.618*	0.274	0.747**	0.262	1	0.53	0.097	0.088
	Sig. (two-tailed)	0.011	0.584	0.186	0.479	0.026	0.074	0.019	0.343	0.002	0.366		0.051	0.742	0.764
Sugar	Pearson correlation	−0.855**	−0.356	−0.169	−0.511	−0.532	−0.752**	0.452	0.45	0.852**	−0.118	0.53	1	−0.016	0.09
	Sig. (two-tailed)	0	0.212	0.565	0.062	0.05	0.002	0.105	0.106	0	0.689	0.051		0.956	0.76
NCSM	Pearson correlation	−0.092	−0.309	0.710**	−0.105	0.361	0.083	−0.515	−0.389	−0.239	0.567*	0.097	−0.016	1	0.844**
	Sig. (two-tailed)	0.754	0.283	0.004	0.72	0.205	0.777	0.06	0.17	0.411	0.035	0.742	0.956		0
Polyphenol	Pearson correlation	−0.176	−0.206	0.819**	−0.285	0.255	0.106	−0.543*	−0.314	−0.179	0.242	0.088	0.09	0.844**	1
	Sig. (two-tailed)	0.547	0.48	0	0.324	0.379	0.717	0.045	0.274	0.54	0.404	0.764	0.76	0	

Bold numbers mean significant difference. *significant correlation (2-tailed) with *P*-value less than 0.05, **significant correlation (2-tailed) with *P*-value less than 0.01, NCSM, nitrogen-containing secondary metabolites.

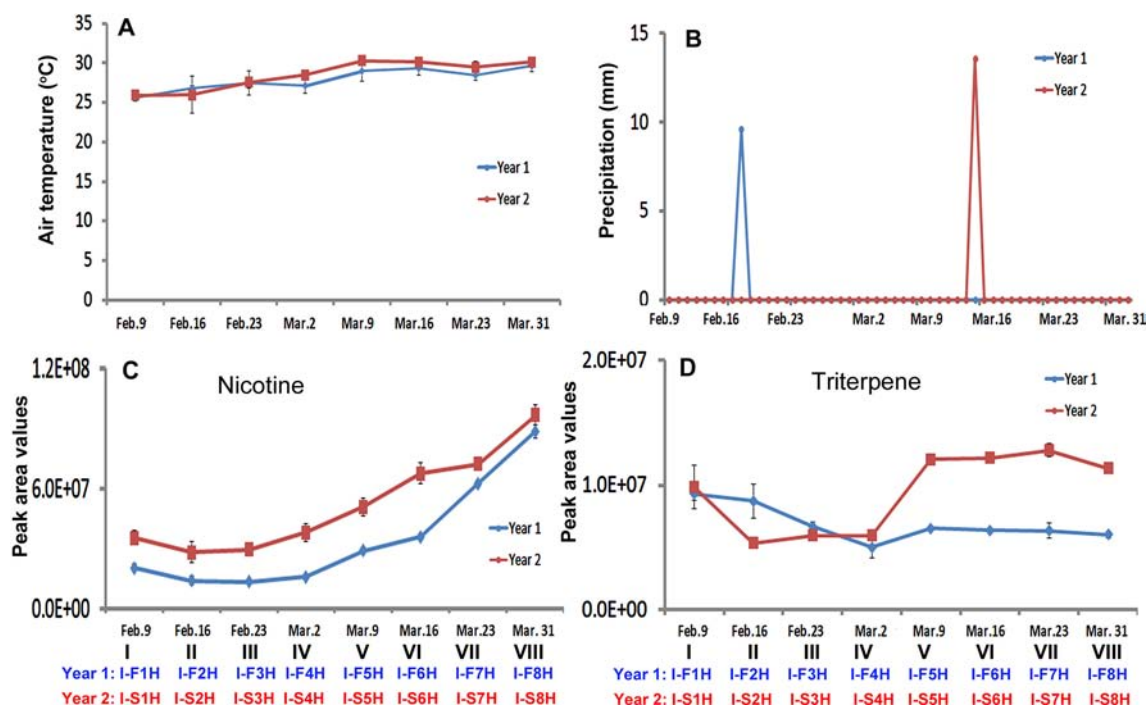


FIGURE 7 | Dynamics of two climate factors and trends of two groups of metabolites in two harvesting seasons in India. **(A–B)**, Dynamics of average temperature **(A)** and precipitation **(B)** values at eight harvesting dates. Each value resulted from averaging daily temperature or precipitation levels in that week of harvest. **(C–D)** Trends of total nicotine levels **(C)** triterpene levels **(D)** in eight groups of samples harvested on the eight harvesting dates.

PCA (Eigen-vector based scaling) was performed to characterize the correlation of 12 groups of metabolites, 2 years, and eight groups of samples. The resulting plot showed that the PC1 and the PC2 accounted for 35.2 and 22.7% of the total variance from two different years (**Figure 8A**). Eight groups of samples (I–VIII) in

the first and second years could be ordinately grouped together, respectively. In comparison, the ordinate association was better in the first year than in the second year. These results indicate differential effects of the two growing years on the profiles of 12 group metabolites.

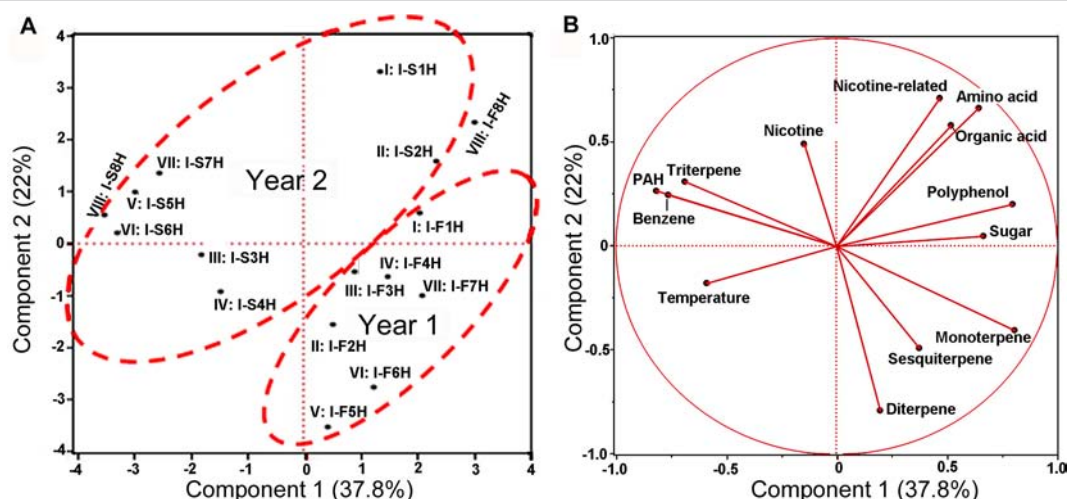


FIGURE 8 | Scaling principal component analysis (PCA) showing associations between metabolite profiles with growing years and two climatic factors in India. This analysis was completed with JUMpro software. **(A)**, A plot of PCA results show an association between two growing years and metabolite profiles in eight groups of samples. **(B)**, A plot of PCA results shows associations between 2-year's temperature and 12 groups of metabolites (including nicotine) as well as correlations among 12 metabolites themselves in 2 years of field growth.

Eight average temperature values, eight average precipitation values, and 12 groups of metabolites were used as variables for PCA. The resulting PCA plot showed potential ordinate relevance between temperature and 12 groups of metabolites (**Figure 8B**). Given that most of precipitation values were zero, precipitation was absent in the plot. This plot showed an opposite ordinate relevance between temperature and sugars, polyphenol, amino acids, organic acids of TCA, and nitrogen-containing secondary metabolites, suggesting a potential negative association. By contrast, this plot showed a close ordinate relevance of temperature with PAH, triterpenes, and benzene in the PC1 axis, suggesting a potential positive association. Pearson analysis was carried out using eight average temperature values, levels of 12 groups of metabolites in 2 years, and eight groups of samples. As described in PCA, precipitation was not included due to almost no rain. A bi-variate Pearson analysis revealed a linear relationship among variables (**Table 2**). The resulting correlation parameters showed that temperature was positively correlated with benzene and PAH (P-values less than 0.05), but negatively correlated with amino acid and organic acid (P-values less than 0.05).

Leaf Nicotine Trends Are Not Altered by Two Growing Locations

Although our experimental design did not specifically target nicotine, we used untargeted data to compare nicotine profiles in leaves from India vs. North Carolina. The rationale is that unlike other metabolites, nicotine is the signature metabolite of tobacco, which is biosynthesized in roots and stored in leaves. The resulting data showed that although the levels of nicotine between 2 years were altered in samples from both India and North Carolina, the profile trends from the bottom to top leaves were similar (**Figures 5C, 7C**). Pearson analysis showed a negative association between nicotine profiles and temperature in North Carolina in 2 years. By contrast, Pearson analysis did not show a significant association between temperature and nicotine profiles in India. In addition, there was not a significant association between precipitation and nicotine profiles in North Carolina. Pearson analysis could not be performed to evaluate the association between precipitation and nicotine profiles in India due to almost no rain in two farming seasons (**Figure 7B**). In summary, these findings show that the trends of nicotine levels from the bottom to top leaves are independent upon local climates and farming locations.

DISCUSSION

Association of Growing Locations and Plant Metabolomes

To date, environmental changes have been reported to highly influence the sensitivity of plants to environments and plant productivity (Aggarwal, 2008; Ahmad et al., 2010; Walter et al., 2010; Gosling et al., 2011). Furthermore, numerous studies have particularly reported that global climate change can highly affect agricultural productivity and crop yield (Walter et al., 2010; Gosling et al., 2011; Bowen and Friel, 2012; Dwivedi et al., 2013). Most of this type of research has focused on single climate factors,

such as temperature in a controlled condition or a specific local environment. A recent study was completed to understand the metabolic responses of rice to temperature changes (Glaubitz et al., 2015). This study used 12 rice cultivars, having different sensitivity to high temperature conditions, to understand their metabolic reactions in vegetative tissues in a given condition. The results showed that the tricarboxylic acid cycle and amino acid biosynthesis in sensitive cultivars were significantly affected by high night temperature and the levels of certain metabolites such as putrescine, spermidine, and spermine were increased in sensitive cultivars (Glaubitz et al., 2015). Another greenhouse study showed that effects of elevated CO₂, elevated temperature, and water deficit alone or in combination differentially affected grape performance in both grape yield and plant growth in a cultivar-dependent manner (Kizildeniz et al., 2015). Drought under an elevated temperature was found to drastically inhibit the vegetative growth with reduced bunch fresh and dry weights of two grape cultivars. Increased carbon dioxide, elevated temperature, and drought were found to reduce the total polyphenolics (Kizildeniz et al., 2015). Similar studies have been completed to understand the effects of single or combined climate factors on other crop growth and production (Sengar et al., 2015). However, studies on plant metabolomes responsive to global climate changes are limited. Moreover, effects of real field climate on crop metabolism in the farming field remains largely open for investigation. In the study reported herein, the objective was to use the globally commercialized K326 tobacco cultivar to understand the effects of real field environment conditions on plant metabolome. We selected two research station fields in two completely different environmental conditions, India and USA. The planting of K326 and field management were exactly followed as per location-specific farming protocols. Using an untargeted metabolomics to profile metabolites we were able to annotate 171 non-polar metabolites from 387 assays and 225 polar metabolites from 366 assays. Further PCA and HCA using these metabolites provided the same result that although plants were grown in two continents, the non-polar and polar metabolomes annotated were primarily grouped and clustered by growing years (**Figures 1, 2, 3, and 4**). These results indicated that the two growing years played a primary role leading to differentiation of metabolomes. Moreover, plants were grown in the same place and managed with the same field protocols during both the years. Both Pearson analysis and PCA revealed positive or negative relevance between temperature and six groups of metabolites in two years in North Carolina (**Table 1**). In addition, precipitation in North Carolina was also significantly associated with triterpenes (**Table 1**). These results suggest that climate changes between 2 years are the main factor associated with the observed plant metabolism differentiation.

Effects of Field Air Temperature on Metabolome

All environmental factors, such as temperature, precipitation, intensity of light, photoperiod, soil condition, wind, pests, diseases, ultraviolet lights, nutrients, soil moisture, and others can significantly affect plant growth and metabolism in the field. Each environmental factor plays a significant role. Therefore, metabolite

TABLE 2 | Data of Pearson correlation analysis for temperature and 12 groups of metabolites in India.

		Temperature	Monoterpene	Sesquiterpene	Diterpene	Triterpene	Benzene	PAH	Nicotine	Amino acid	Organic acid	Sugar	NCSM	Polyphenol
Temperature	Pearson correlation	1	−0.3	0.186	−0.1	0.38	0.613*	0.499*	0.12	−0.551*	−0.588*	−0.024	−0.207	−0.236
Sig. (two-tailed)		0.26	0.49	0.714	0.147	0.012	0.049	0.659	0.027	0.017	0.931	0.442	0.379	
Monoterpene	Pearson correlation	−0.3	1	0.559*	0.488	−0.477	−0.580*	−0.642**	−0.523*	0.317	0.145	0.633**	0.158	0.672**
Sig. (two-tailed)		0.26	0.024	0.055	0.062	0.018	0.007	0.038	0.231	0.591	0.009	0.56	0.004	
Sesquiterpene	Pearson correlation	0.186	0.559*	1	0.528*	−0.23	−0.245	−0.203	−0.152	−0.05	−0.155	0.581*	0.108	0.211
Sig. (two-tailed)		0.024	0.035	0.035	0.391	0.361	0.451	0.575	0.856	0.567	0.018	0.691	0.432	
Diterpene	Pearson correlation	−0.1	0.488	0.528*	1	−0.309	−0.305	−0.313	−0.425	−0.317	−0.078	−0.063	−0.371	−0.222
Sig. (two-tailed)		0.714	0.055	0.035	0.244	0.251	0.238	0.101	0.232	0.773	0.818	0.158	0.41	
Triterpene	Pearson correlation	0.38	−0.477	−0.23	−0.309	1	0.802**	0.811**	−0.079	−0.068	−0.138	−0.251	−0.067	−0.454
Sig. (two-tailed)		0.147	0.062	0.391	0.244	0	0	0.772	0.803	0.61	0.348	0.805	0.077	
Benzene	Pearson correlation	0.613*	−0.580*	−0.245	−0.305	0.802**	1	0.803**	−0.046	−0.199	−0.207	−0.232	−0.167	−0.422
Sig. (two-tailed)		0.012	0.018	0.361	0.251	0	0	0.867	0.46	0.443	0.386	0.536	0.103	
PAH	Pearson correlation	0.499*	−0.642**	−0.203	−0.313	0.811**	0.803**	1	0.131	−0.292	−0.317	−0.351	−0.034	−0.575*
Sig. (two-tailed)		0.049	0.007	0.451	0.238	0	0	0.627	0.273	0.231	0.182	0.902	0.02	
Nicotine	Pearson correlation	0.12	−0.523*	−0.152	−0.425	−0.079	−0.046	0.131	1	−0.021	0.087	−0.151	0.425	−0.062
Sig. (two-tailed)		0.659	0.575	0.101	0.772	0.867	0.627		0.939	0.749	0.578	0.101	0.821	
Amino acid	Pearson correlation	−0.551*	0.317	−0.05	−0.317	−0.068	−0.199	−0.292	−0.021	1	0.805**	0.504*	0.759**	0.595*
Sig. (two-tailed)		0.027	0.231	0.856	0.232	0.803	0.46	0.273	0.939	0	0.046	0.001	0.015	
Organic acid	Pearson correlation	−0.588*	0.145	−0.155	−0.078	−0.138	−0.207	−0.317	0.087	0.805**	1	0.149	0.603*	0.315
Sig. (two-tailed)		0.017	0.591	0.567	0.773	0.61	0.443	0.231	0.749	0	0.582	0.013	0.235	
Sugar	Pearson correlation	−0.024	0.633**	0.581*	−0.063	−0.251	−0.232	−0.351	−0.151	0.504*	0.149	1	0.426	0.780**
Sig. (two-tailed)		0.931	0.009	0.018	0.818	0.348	0.386	0.182	0.578	0.046	0.582	0.1	0	
NCSM	Pearson correlation	−0.207	0.158	0.108	−0.371	−0.067	−0.167	−0.034	0.425	0.759**	0.603*	0.426	1	0.49
Sig. (two-tailed)		0.442	0.56	0.691	0.158	0.805	0.536	0.902	0.101	0.001	0.013	0.1	0.054	
Polyphenol	Pearson correlation	−0.236	0.672**	0.211	−0.222	−0.454	−0.422	−0.575*	−0.062	0.595*	0.315	0.780**	0.49	1
Sig. (two-tailed)		0.379	0.004	0.432	0.41	0.077	0.103	0.02	0.821	0.015	0.235	0	0.054	

Bold numbers mean significant difference. *significant correlation (2-tailed) with P-value less than 0.05, **significant correlation (2-tailed) with P-value less than 0.01, NCSM, nitrogen-containing secondary metabolites.

profile data obtained from 2 years of investigation actually resulted from the comprehensive effect of all field factors in two growing locations. We understood that data for all environmental factors were essential to understand their effects on tobacco metabolism. Although it was difficult to track accurate data of all factors, we could obtain accurate values (hourly and daily) for two weather factors from local weather stations, air temperature and precipitation. Therefore, both air temperature and precipitation of two growing seasons were used to understand their association with metabolite profiles.

Temperature has been reported a primary factor that can affect plant metabolism. Single temperature factor in the control condition has been showed to affect plant metabolism in different ways. A genome-wide study of *Arabidopsis* indicated metabolic network changes caused by different temperature conditions (Topfer et al., 2013). The increase of certain amino acids during seed imbibition of *Ricinus communis* was reported to be responsive to temperature increase to 35°C (Ribeiro et al., 2015). Furthermore, the reprogramming of the metabolome was reported to occur in temperature stress. For example, the central carbohydrate metabolism can be regulated by the temperature-stress (Guy et al., 2008). In our study, we used Pearson analysis to characterize association of the field air temperatures in two different geographical areas with leaf metabolome composition. To understand the potential association of temperature and metabolome, we selected nicotine and 11 groups of metabolites (Figures 5 and 7, Supplementary Figures 6 and 7, and Supplementary Table 9). The 11 groups included both plant primary and secondary metabolites (Supplementary Table 9). The total level of each group at different harvest date was summed from individual metabolites, and then used for characterization of their level trends, PCA, and Pearson analysis. The resulting data allowed evaluating association of metabolite groups with field air temperature. We selected nicotine as a representative metabolite, because this health-associated alkaloid is the signature metabolite of tobacco. The resulting dynamic trend data developed from each harvest time, PCA, and Pearson analysis simultaneously showed that the levels of nicotine were negatively associated with field air temperatures in 2 years of growth in North Carolina (Figures 5A, C, 6B, Supplementary Figure 6D, and Table 1). By contrast, although the trends of temperature and nicotine levels in India (Figures 7A, C) were similar to those in North Carolina, the results from PCA and Pearson analysis showed a potentially partial association (Figure 8B and Table 2). This observation was consistent with our previous data reporting nicotine differentiation in tobacco samples from India, USA, and Brazil (Ma et al., 2013).

In addition, we observed the associations between temperature and other metabolites. It was interesting to note that the total level of organic acids of TCA cycle (Supplementary Table 8) was negatively associated with temperature in both North Carolina and India (Tables 1 and 2). One of reduced organic acids is malonic acid, an intermediate in the TCA cycle (Ferne et al., 2004). The reduction of this acid and other acids of TCA by high temperature was also observed in soybean (Sicher, 2013). These data suggest that the accumulation of organic acids of TCA is associated with the field air temperature. Moreover, we observed the different responses of metabolites to temperatures in two locations. In North Carolina,

temperature was positively associated with sesquiterpenes, diterpenes, and triterpenes, but negatively associated with sugars (Table 1). In India, temperature was positively associated with PAH and benzene but was negatively associated with amino acids. These responsive differences likely result from multiple environmental factors in the field. Experimental evidence has showed that a combination of environmental factors such as heat shock and drought can lead to multiple alterations in plant metabolism, as in photosynthesis and enzyme activity (Rizhsky et al., 2002). Extreme temperature (together with water deficit) and high solar radiation were reported to strongly affect grapevine growth, which further led to negative effects on fruit and wine quality (Teixeira et al., 2014). A field study in China revealed that different field conditions also affected the profiles of polar metabolites in tobacco leaves such as flavonoids (Li et al., 2014). In 2 years of our field experiments, it rained only once during each year in India. We hypothesize that this extreme precipitation event can affect effects of temperature on plant metabolism in India. In summary, temperature conditions play a primary role in regulating plant metabolism in the field. We further believe that as more field experiments will be performed, continuous documentation of data will enhance the understanding of combined effects of temperature and other factors on plant metabolism.

Effects of Precipitation on Metabolome

Global changes in precipitation regimes have been continuously documented (Keller, 2007; Schnoor, 2007). Those precipitation changes are associated with different factors, such as rising temperature (Le Roux et al., 2013; Vlam et al., 2014). Variation in precipitation patterns such as water deficit (drought) have been reported to lead to lower the photosynthetic rates and plant growth as well as other physiological features (Munne-Bosch and Penuelas, 2003; Tezara et al., 2003). In comparison, studies on effects of variation in precipitation patterns such as drought on metabolites accumulation remain open for study. In this study, we recorded daily precipitation and calculated average precipitation during the growing seasons in two continuous years. The weather was dry in India during these 2 years, wherein the field received raining only once during the growing season in each year (Figure 7B). Consequently, we could only analyze relevance of precipitation and metabolomes for North Carolina samples. The weekly average and daily precipitation values were used for PCA and Pearson analysis, respectively. The resulting PCA data showed a positive relevance among average precipitation, temperature, triterpenes, and sesquiterpenes (Figure 6B). The resulting Pearson analysis data showed a significant association between daily precipitation and triterpenes. These data provide evidence that precipitation can affect plant metabolism, although it is hard to predict precipitation year by year.

Leaf Nicotine Trends Independent of Environmental Factors

Nicotine is the main signature metabolite of tobacco. It is formed in the roots, transported to above ground tissues, and stored in the leaves. Our study reveals an interesting metabolic independence

of nicotine level trends upon climate from 2 years of studies. This independent feature is that although Pearson analysis and PCA showed positive and negative associations between nicotine and temperature in North Carolina and India, respectively, the trend of nicotine levels was not altered in 2 years of experiments. In two fields, the bottom leaves had lower nicotine levels, while the top leaves had the higher nicotine levels (Figures 6C and 7C). In contrast, the level trends of 11 other groups of metabolites in different groups of samples were dynamic in tobacco leaves from both North Carolina and India. These interesting results indicate that the trend of nicotine levels is stably controlled by tobacco plants. The mechanism behind this trend stability of nicotine remain uninvestigated, although its biosynthetic pathway has been studied intensively (Xie and Fan, 2016; Kajikawa et al., 2017). Furthermore, it has been vastly documented that nicotine is exclusively biosynthesized in roots and then transported to leaves for storage. In our study, we used standardized tobacco farming protocols to grow plants in the field. Therefore, the use of fertilizer and the time of fertilization were consistent in 2 years. Based on these observations we hypothesize that although temperature and other environmental factors can affect nicotine levels in leaves, weather factors might have limited effects on the trend of nicotine levels controlled by the root-specific biosynthesis. This finding shows that it is interesting to understand effects of environmental factors on the formation of root-specific metabolites in the future.

CONCLUSION REMARKS

Herein, we integrate unbiased un-targeted metabolomics with standard farming practices followed for commercial cultivation of tobacco to understand the formation of root-specific effects of field environments on plant metabolism. From a large number of peaks detected, we could annotate 171 non-polar and 225 polar metabolites and further grouped them into different classes. Nicotine, the main tobacco alkaloid, is one of the non-polar metabolites and used as signature metabolite in our data analysis. PCA and HCA characterized that two different years and two different geographical locations played primary and secondary roles in controlling metabolite complexity. Field air temperature was characterized to be one of the main factors that was associated

with profiles of several main groups of metabolites. We further used nicotine as a signature metabolite to indicate that its profile trend in different leaves controlled by root-specific biosynthesis is independent of climate factors. This pilot study provides useful findings to enhance the understanding of the effects of two diverse climate factors on plant metabolism across two continents.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

D-YX and SG conceived this research and designed all experiments in North Carolina and India, respectively. D-YX and SG also participated in sample collection and collected weather data in North Carolina and India, respectively. D-YX participated in data analysis and figure preparation, and drafted and finalized this manuscript. D-MM managed plant growth in field in North Carolina, collected samples, performed metabolic extraction and GC-MS analysis and other lab experiments, analyzed data, and prepared figures and manuscript. CH participated in field design and management of plant growth, involved in sampling, and metabolite extraction. RM participated in field design and management and sampling in India.

ACKNOWLEDGMENTS

This project was supported by Life Sciences and Technology Centre, Bangalore of ITC Limited, India.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.01370/full#supplementary-material>

REFERENCES

- Abberton, M., Batley, J., Bentley, A., Bryant, J., Cai, H. W., Cockram, J., et al. (2016). Global agricultural intensification during climate change: a role for genomics. *Plant Biotechnol. J.* 14, 1095–1098. doi: 10.1111/pbi.12467
- Adams, R. M., Hurd, B. H., Lenhart, S., and Leary, N. (1998). Effects of global climate change on agriculture: an interpretative review. *Clim. Res.* 11, 19–30. doi: 10.3354/cr011019
- Aggarwal, P. K. (2008). Global climate change and Indian agriculture: impacts, adaptation and mitigation. *Indian J. Agric. Sci.* 78, 911–919.
- Ahmad, A., Diwan, H., and Abrol, Y. P. (2010). *Global Climate Change, Stress and Plant Productivity*. New York: Springer. doi: 10.1007/978-90-481-3112-9_23
- Almaarri, K., Alamir, L., Junaid, Y., and Xie, D.-Y. (2010). Volatile compounds from leaf extracts of *Juniperus excelsa* growing in Syria via gas chromatography mass spectrometry. *Analytical Methods* 2, 673–677. doi: 10.1039/b9ay00256a
- Bowen, K. J., and Friel, S. (2012). Climate change adaptation: Where does global health fit in the agenda? *Global Health* 8 (1), 10. doi: 10.1186/1744-8603-8-10
- Chauhan, B. S., Prabhjyot, K., Mahajan, G., Randhawa, R. K., Singh, H., and Kang, M. S. (2014). “Global warming and its possible impact on agriculture in India,” in *Advances in Agronomy*, vol. 123. Ed. Spark, D. L. (San Diego: Elsevier Academic Press Inc), 65–121. doi: 10.1016/B978-0-12-420225-2.00002-9
- Clayton, S., Devine-Wright, P., Stern, P. C., Whitmarsh, L., Carrico, A., Steg, L., et al. (2015). Psychological research and global climate change. *Nat. Clim. Chang.* 5, 640–646. doi: 10.1038/nclimate2622
- Duan, L. X., Chen, T. L., Li, M., Chen, M., Zhou, Y. Q., Cui, G. H., et al. (2012). Use of the metabolomics approach to characterize Chinese medicinal material Huangqi. *Mol. Plant* 5, 376–386. doi: 10.1093/mp/sss093
- Dwivedi, S., Sahrawat, K., Upadhyaya, H., and Ortiz, R. (2013). “Food, Nutrition and Agrobiodiversity Under Global Climate Change,” in *Advances in Agronomy*, vol. 120. Ed. Sparks, D. L. (San Diego: Elsevier Academic Press Inc), 1–128. doi: 10.1016/B978-0-12-407686-0.00001-4

- Eriksen, M., Mackay, J., Schluger, N., Gomeshtapeh, F. I., and Drope, J. (2015). *The Tobacco Atlas*. Atlanta, Georgia 30303 USA: The American Cancer Society, Inc.
- Fernie, A. R., Carrari, F., and Sweetlove, L. J. (2004). Respiratory metabolism: glycolysis, the TCA cycle and mitochondrial electron transport. *Curr. Opin. Plant Biol.* 7, 254–261. doi: 10.1016/j.pbi.2004.03.007
- Glaubitx, U., Erban, A., Kopka, J., Hinch, D. K., and Zuther, E. (2015). High night temperature strongly impacts TCA cycle, amino acid and polyamine biosynthetic pathways in rice in a sensitivity-dependent manner. *J. Exp. Bot.* 66, 6385–6397. doi: 10.1093/jxb/erv352
- Gosling, S. N., Warren, R., Arnell, N. W., Good, P., Caesar, J., Bernie, D., et al. (2011). A review of recent developments in climate change science. Part II: The global-scale impacts of climate change. *Prog. Phys. Geogr.* 35, 443–464. doi: 10.1177/0309133311407650
- Guy, C., Kaplan, F., Kopka, J., Selbig, J., and Hinch, D. K. (2008). Metabolomics of temperature stress. *Physiol. Plant.* 132, 220–235. doi: 10.1111/j.1399-3054.2007.00999.x
- Hauser, D. D. W., Tobin, E. D., Feifel, K. M., Shah, V., and Pietri, D. M. (2016). Disciplinary reporting affects the interpretation of climate change impacts in global oceans. *Glob. Change Biol.* 22, 25–43. doi: 10.1111/gcb.12978
- Jones, N. (2013). Climate assessments: 25 years of the IPCC. *Nature* 501, 298–299. doi: 10.1038/501298a
- Kajikawa, M., Sierro, N., Kawaguchi, H., Bakaher, N., Ivanov, N. V., Hashimoto, T., et al. (2017). Genomic insights into the evolution of the nicotine biosynthesis pathway in tobacco. *Plant Physiol.* 174, 999–1011. doi: 10.1104/pp.17.00070
- Keller, C. F. (2007). Global warming 2007. An update to global warming: the balance of evidence and its policy implications. *Sci. World J.* 7, 381–399. doi: 10.1100/tsw.2007.91
- Kintisch, E. (2013). Climate science. For researchers, IPCC leaves a deep impression. *Science* 342, 24. doi: 10.1126/science.342.6154.24
- Kizildeniz, T., Mekni, I., Santesteban, H., Pascual, I., Morales, F., and Irigoyen, J. J. (2015). Effects of climate change including elevated CO₂ concentration, temperature and water deficit on growth, water status, and yield quality of grapevine (*Vitis vinifera* L.) cultivars. *Agric. Water Manage.* 159, 155–164. doi: 10.1016/j.agwat.2015.06.015
- Kjellstrom, T., Briggs, D., Freyberg, C., Lemke, B., Otto, M., and Hyatt, O. (2016). “Heat, human performance, and occupational health: a key issue for the assessment of global climate change impacts,” in *Annual Review of Public Health*, vol. 37. Ed. (Palo Alto: Annual Reviews), 97–112. doi: 10.1146/annurev-publhealth-032315-021740
- Kollah, B., Patra, A. K., and Mohanty, S. R. (2016). Aquatic microphylla Azolla: a perspective paradigm for sustainable agriculture, environment and global climate change. *Environ. Sci. Pollut. Res.* 23, 4358–4369. doi: 10.1007/s11356-015-5857-9
- Le Roux, P. C., Aalto, J., and Luoto, M. (2013). Soil moisture's underestimated role in climate change impact modelling in low-energy systems. *Glob. Change Biol.* 19, 2965–2975. doi: 10.1111/gcb.12286
- Li, L. L., Zhao, C. X., Chang, Y. W., Lu, X., Zhang, J. J., Zhao, Y. N., et al. (2014). Metabolomics study of cured tobacco using liquid chromatography with mass spectrometry: method development and its application in investigating the chemical differences of tobacco from three growing regions. *J. Sep. Sci.* 37, 1067–1074. doi: 10.1002/jssc.201301138
- Ma, D.-M., Gandra, S. V., Sharma, N., and Xie, D.-Y. (2013). Integration of GC-MS based non-targeted metabolic profiling with headspace solid phase microextraction enhances the understanding of volatile differentiation in tobacco leaves from North Carolina, India and Brazil. *Am. J. Plant. Sci.* 3, 1759–1769. doi: 10.4236/ajps.2012.312215
- Ma, D. M., Wang, Z., Wang, L., Alejos-Gonzales, F., Sun, M. A., and Xie, D. Y. (2015). A Genome-Wide Scenario of Terpene Pathways in Self-pollinated *Artemisia annua*. *Mol. Plant.* 11, 1580–1598. doi: 10.1016/j.molp.2015.07.004
- Munne-Bosch, S., and Penuelas, J. (2003). Photo- and antioxidative protection during summer leaf senescence in *Pistacia lentiscus* L. grown under Mediterranean field conditions. *Ann. Bot.* 92, 385–391. doi: 10.1093/aob/mcg152
- Park, C. E., Jeong, S. J., Ho, C. H., and Kim, J. (2015). Regional variations in potential plant habitat changes in response to multiple global warming scenarios. *J. Clim.* 28, 2884–2899. doi: 10.1175/JCLI-D-13-00753.1
- Patz, J. A., Epstein, P. R., Burke, T. A., and Balbus, J. M. (1996). Global climate change and emerging infectious diseases. *JAMA-J Am. Med. Assoc.* 275, 217–223. doi: 10.1001/jama.1996.03530270057032
- Polley, H. W., Briske, D. D., Morgan, J. A., Wolter, K., Bailey, D. W., and Brown, J. R. (2013). Climate change and North American rangelands: Trends, Projections, and Implications. *Rangel. Ecol. Manag.* 66, 493–511. doi: 10.2111/REM-D-12-00068.1
- Ribeiro, P. R., Mudde, E., Fernandez, L. G., De Castro, R. D., Ligterink, W., and Hilhorst, H. W. M. (2015). Metabolite profiling of the oilseed crop *Ricinus communis* during early seed imbibition reveals a specific metabolic signature in response to temperature. *Ind. Crop. Prod.* 67, 305–309. doi: 10.1016/j.indcrop.2015.01.067
- Rizhsky, L., Liang, H. J., and Mittler, R. (2002). The combined effect of drought stress and heat shock on gene expression in tobacco. *Plant. Physiol.* 130, 1143–1151. doi: 10.1104/pp.006858
- Rodgman, A., and Perfetti, T. A. (2009). *The Chemical Components of Tobacco and Tobacco Smoke*. Boca Raton: CRC Press, Taylor & Francis Group. doi: 10.1201/9781420078848
- Rom, W. N., and Pinkerton, K. E. (2014). *Introduction: Consequences of Global Warming to the Public's Health*. Totowa: Humana Press Inc. doi: 10.1007/978-1-4614-8417-2_1
- Schmitz, O. J. (2013). “Global climate change and the evolutionary ecology of ecosystem functioning,” in *Climate Change and Species Interactions: Ways Forward*. Eds. Angert, A. L., Ladeau, S. L., and Ostfeld, R. S. (Oxford: Blackwell Science Publ), 61–72.
- Schnoor, J. L. (2007). The IPCC fourth assessment. *Environ. Sci. Technol.* 41, 1503. doi: 10.1021/es072475x
- Sengar, K., Sengar, R. S., Lal, K., and Rao, V. P. (2015). *Climate change effect on sugarcane productivity. Climate Change Effect on Crop Productivity*. R. S. Sengar and K. Sengar. Boca Raton, FL, USA: CRC Press-Taylor & Francis Group. 177–186.
- Sicher, R. (2013). Combined effects of CO₂ enrichment and elevated growth temperatures on metabolites in soybean leaflets: evidence for dynamic changes of TCA cycle intermediates. *Planta* 238, 369–380. doi: 10.1007/s00425-013-1899-8
- Sierro, N., Battey, J. N. D., Ouadi, S., Bakaher, N., Bovet, L., Willig, A., et al. (2014). The tobacco genome sequence and its comparison with those of tomato and potato. *Nat. Commun.* 5, 3833. doi: 10.1038/ncomms4833
- Taylor, S., and Kumar, L. (2016). Global climate change impacts on pacific islands terrestrial biodiversity: a review. *Trop. Conserv. Sci.* 9, 203–223. doi: 10.1177/194008291600900111
- Teixeira, A., Martins, V., Noronha, H., Eiras-Dias, J., and Geros, H. (2014). The first insight into the metabolite profiling of grapes from three *Vitis vinifera* L. cultivars of two controlled appellation (DOC) regions. *Int. J. Mol. Sci.* 15, 4237–4254. doi: 10.3390/ijms15034237
- Tezara, W., Martinev, D., Rengifo, E., and Herrera, A. (2003). Photosynthetic responses of the tropical spiny shrub *Lycium nodosum* (Solanaceae) to drought, soil salinity and saline spray. *Ann. Bot.* 92, 757–765. doi: 10.1093/aob/mcg199
- Topfer, N., Caldana, C., Grimbs, S., Willmitzer, L., Fernie, A. R., and Nikoloski, Z. (2013). Integration of genome-scale modeling and transcript profiling reveals metabolic pathways underlying light and temperature acclimation in *Arabidopsis*. *Plant. Cell.* 25, 1197–1211. doi: 10.1105/tpc.112.108852
- Vlam, M., Baker, P. J., Bunyavechewin, S., and Zuidema, P. A. (2014). Temperature and rainfall strongly drive temporal growth variation in Asian tropical forest trees. *Oecologia* 174, 1449–1461. doi: 10.1007/s00442-013-2846-x
- Walter, L. C., Streck, N. A., Rosa, H. T., and Kruger, C. (2010). Climate change and its effects on rice. *Cienc. Rural* 40, 2411–2418. doi: 10.1590/S0103-84782010001100028
- Xie, J. H., and Fan, L. J. (2016). Nicotine biosynthesis is regulated by two more layers: small and long non-protein-coding RNAs. *Plant Signal Behav.* 11, UNSP e1184811. doi: 10.1080/15592324.2016.1184811

Conflict of Interest: The authors declare that the research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Ma, Gandra, Manoharlal, La Hovary and Xie. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Contribution of Functional Divergence Through Copy Number Variations to the Inter-Species and Intra-Species Diversity in Specialized Metabolites

Kazumasa Shirai and Kousuke Hanada*

Department of Bioscience and Bioinformatics, Kyushu Institute of Technology, Fukuoka, Japan

OPEN ACCESS

Edited by:

Takayuki Tohge,
Nara Institute of Science and
Technology (NAIST), Japan

Reviewed by:

Dirk Walther,
Max Planck Institute of Molecular
Plant Physiology, Germany
Qing Liu,
Commonwealth Scientific and
Industrial Research Organisation
(CSIRO), Australia

*Correspondence:

Kousuke Hanada
kohanada@bio.kyutech.ac.jp

Specialty section:

This article was submitted to
Plant Metabolism
and Chemodiversity,
a section of the journal
Frontiers in Plant Science

Received: 18 July 2019

Accepted: 08 November 2019

Published: 26 November 2019

Citation:

Shirai K and Hanada K (2019)
Contribution of Functional Divergence
Through Copy Number Variations to
the Inter-Species and Intra-Species
Diversity in Specialized Metabolites.
Front. Plant Sci. 10:1567.
doi: 10.3389/fpls.2019.01567

There is considerable diversity in the specialized metabolites within a single plant species (intra-species) and among different plant species (inter-species). The functional divergence associated with gene duplications largely contributes to the inter-species diversity in the specialized metabolites, whereas the intra-species diversity is due to gene dosage changes *via* gene duplications [i.e., copy number variants (CNVs)] at the intra-species level of evolution. This is because CNVs are thought to undergo associated with less functional divergence at the intra-species level of evolution. However, functional divergence caused by CNVs may induce specialized metabolite diversity at the intra-species and inter-species levels of evolution. We herein discuss the functional divergence of CNVs in metabolic quantitative trait genes (mQTGs). We focused on 5,654 previously identified mQTGs in 270 *Arabidopsis thaliana* accessions. The ratio of nonsynonymous to synonymous variations tends to be higher for mQTGs with CNVs than for mQTGs without CNVs within *A. thaliana* accessions, suggesting that CNVs are responsible for the functional divergence among mQTGs at the intra-species level of evolution. To evaluate the contribution of CNVs to inter-species diversity, we calculated the ratio of nonsynonymous to synonymous substitutions in the *Arabidopsis* lineage. The ratio tends to be higher for the mQTGs with CNVs than for the mQTGs without CNVs. Additionally, we determined that mQTGs with CNVs are subject to positive selection in the *Arabidopsis* lineage. Our data suggest that CNVs are closely related to functional divergence contributing to adaptations *via* the production of diverse specialized metabolites at the intra-species and inter-species levels of evolution.

Keywords: specialized metabolite, adaptation, *Arabidopsis*, copy number variant, gene duplication

INTRODUCTION

Plants produce various specialized metabolites, the diversity of which is closely related to adaptive evolution (Pichersky and Lewinsohn, 2011). Specialized metabolites vary among different species as well as within single species (Chan et al., 2010; Weigel, 2012; Carreno-Quintero et al., 2013; Alseekh et al., 2015; Matsuda et al., 2015). The diversity of the specialized metabolites resulted from gene duplications among various plant species. We previously revealed that copy number variants

(CNVs) derived from gene duplications are associated with specialized metabolites (Shirai et al., 2017).

Gene duplications contribute to the diversity in specialized metabolites because of two possible effects. The first effect is functional divergence. After gene duplication events, the copied genes tend to accumulate nonsynonymous mutations because of relaxed selection pressures (Scannell and Wolfe, 2008). Consequently, the copied genes induce functional divergence (Ohno, 1970), ultimately leading to the variability in the specialized metabolites among various plant species (Hanada et al., 2008; Kliebenstein and Osbourn, 2012; Panchy et al., 2016). The second effect involves gene dosage changes. Specifically, gene duplications increase gene dosage (Ohno, 1970). In particular, CNVs are believed to be the main cause of intra-species gene dosage changes (Zmienko et al., 2014). There is experimental evidence that the abundance of specialized metabolites within a single species is critically controlled by altered gene dosages due to CNVs (Kliebenstein, 2001). However, it remains unclear whether CNVs associated with specialized metabolites tend to induce functional divergence at the genomic scale.

We herein discuss the functional divergence of CNVs associated with specialized metabolites. For this discussion, we performed additional analyses involving our previously published data. On the basis of the analyses, we propose that CNVs induce functional divergence that generates various specialized metabolites during the evolution of *A. thaliana*.

Functional Divergence of CNVs at the Intra-Species Level of Evolution

It is believed that CNVs mainly cause quantitative changes rather than qualitative changes (Zmienko et al., 2014), likely because of an insufficient amount of time for CNVs to accumulate nonsynonymous mutations leading to the diversity in specialized metabolites. However, several studies have identified a few nonsynonymous mutations responsible for the functional divergence of genes related to specialized metabolites (Chye et al., 2000; Yu et al., 2015; Bunsupa et al., 2016). These reports suggest CNVs may induce functional divergence.

To examine the functional divergence of duplicated genes, the selection pressure based on the ratio between the nonsynonymous mutation/substitution rate and the synonymous mutation/substitution rate is useful (Hanada et al., 2009). High and low selection pressures are associated with functional divergence and constraint, respectively. Therefore, we estimated the $dN_{\text{SNP}}/dS_{\text{SNP}}$ ratio, which is the ratio between the number of nonsynonymous mutations relative to the number of nonsynonymous sites (dN_{SNP}) and the number of synonymous mutations relative to the number of synonymous sites (dS_{SNP}) (Nei and Gojobori, 1986; Hanada et al., 2009), for 27,130 annotated protein-coding genes in 270 *A. thaliana* accessions.

The single nucleotide polymorphism (SNP) data examined in this study were compiled from 270 *A. thaliana* accessions analyzed

in several studies (<http://1001genomes.org>, 1001 Genomes; Mouille et al., 2006; Cao et al., 2011; Gan et al., 2011; Schmitz et al., 2013; Shirai et al., 2017). A total of 7,624,270 SNPs were included. For each accession, nonsynonymous and synonymous variations were annotated according to the TAIR10 database with the SnpEff program (<https://www.arabidopsis.org>; Cingolani et al., 2012). Of the 7,624,270 SNPs, 1,330,920 were located in 27,130 annotated protein-coding genes in the reference *A. thaliana* genome. For each of the 270 accessions, the 1,330,920 SNPs were classified as 733,796 nonsynonymous and 597,124 synonymous mutations in the 27,130 coding sequences. There was an average of 27 nonsynonymous and 22 synonymous mutations in each of the 27,130 coding sequences. Because the number of nonsynonymous and synonymous sites in codons varied, we calculated the number of synonymous and nonsynonymous sites in all 27,130 coding sequences with scripts that we developed following the Nei-Gojobori method (Nei and Gojobori, 1986; Hanada et al., 2009). The 27,130 genes were classified as metabolic quantitative trait genes (mQTGs) and mQTGs with CNVs. We previously predicted 5,654 mQTGs for 1,335 specialized metabolites in *A. thaliana* (Shirai et al., 2017). In that study, mQTGs were detected by combining a genome-wide association study (GWAS) and a metabolite-transcriptome correlation analysis (MTCA). This method enabled the prediction of mQTGs with a lower false positive rate than that of the general GWAS method. Genes with CNVs were previously detected by comparing genomic read counts among *A. thaliana* accessions (Gan et al., 2011). Of the 27,130 genes, 929 were predicted as genes with CNVs ($P < 0.05$).

To assess whether the functional divergence of CNVs is associated with the diversity in specialized metabolites, we compared the $dN_{\text{SNP}}/dS_{\text{SNP}}$ ratios among mQTGs, mQTGs with CNVs, and randomly selected genes (**Figure 1A** and **Supplementary Table S1**). The $dN_{\text{SNP}}/dS_{\text{SNP}}$ ratios were significantly higher for the mQTGs and mQTGs with CNVs than for the randomly selected genes (Wilcoxon rank sum test: $P < 0.001$; **Figure 1A**). Moreover, the ratios of mQTGs with CNVs were also significantly higher than the ratios of mQTGs (Wilcoxon rank sum test: $P < 0.001$; **Figure 1A**). These results imply that nonsynonymous variations tend to accumulate in mQTGs more frequently than in genes not associated with specialized metabolites. Specifically, mutations that alter the amino acid sequence accumulated in mQTGs with CNVs at a higher rate than in mQTGs without CNVs. These findings suggest that the diversity in specialized metabolites due to CNVs is the result of the functional divergence of mQTGs in addition to gene dosage changes at the genomic scale. Additionally, mQTGs with CNVs tended to be associated with a larger number of specialized metabolites than mQTGs without CNVs (Wilcoxon rank sum test: $P = 1.92 \times 10^{-3}$; **Supplementary Figure S1**), implying that the functional divergence derived from CNVs enhances the divergence of specialized metabolites.

It was unclear whether CNVs induce functional divergence for mQTGs only or for other genes as well. Therefore, we compared the $dN_{\text{SNP}}/dS_{\text{SNP}}$ ratios of randomly selected genes and non-mQTGs with CNVs. The $dN_{\text{SNP}}/dS_{\text{SNP}}$ ratios were significantly higher for the non-mQTGs with CNVs than for the randomly selected genes (Wilcoxon rank sum test: $P < 2.2 \times$

Abbreviations: CNV, copy number variant; mQTG, metabolic quantitative trait gene; NI, neutrality index; PacBio, Pacific Bioscience; SNP, single nucleotide polymorphism.

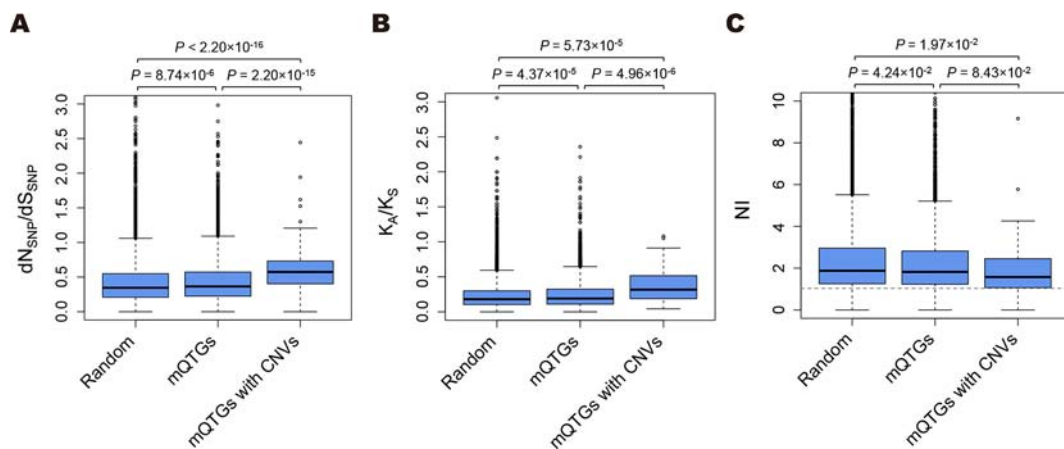


FIGURE 1 | Comparison of the functional divergence and the selection pressure in *Arabidopsis*. **(A)** Box plots present the $dN_{\text{SNP}}/dS_{\text{SNP}}$ ratios of *Arabidopsis thaliana* accessions. **(B)** Box plots present the K_A/K_S ratios between *A. thaliana* and *Arabidopsis lyrata*. **(C)** Box plots present the NIs between *A. thaliana* and *A. lyrata*. Random, 10,000 randomly selected genes; mQTGs, metabolic quantitative trait genes; mQTGs with CNVs, metabolic quantitative trait genes with copy number variants. In each box plot, the box represents the 25%–75% range, the middle line represents the median, the dotted line represents the 1%–99% range, and the outer circles are outliers. *P* values were calculated with the Wilcoxon rank sum test. The horizontal dotted line represents $NI = 1.0$.

10^{-16} ; **Supplementary Figure S2**). Thus, CNVs are generally responsible for the functional divergence of genes at the intra-species level of evolution.

Functional Divergence of CNVs at the Inter-Species Level of Evolution

It was recently reported that CNVs are associated with various phenotypic differences within a plant species (Lye and Purugganan, 2019). By contrast, the contribution of CNVs to inter-species diversity remains relatively unknown in plants.

The $dN_{\text{SNP}}/dS_{\text{SNP}}$ ratio indicates the intra-species level of evolution. Therefore, to characterize the functional divergence of CNVs at the inter-species level of evolution, we examined the K_A/K_S ratio, which is the ratio between the number of nonsynonymous substitutions relative to the number of nonsynonymous sites (K_A) and the number of synonymous substitutions relative to the number of synonymous sites (K_S). The K_A/K_S ratio was estimated for 20,498 orthologs between *A. thaliana* and *Arabidopsis lyrata*. These orthologs were detected based on the reciprocal best hit (E-value $< 1.0 \times 10^{-3}$ and coverage $> 90\%$) of a BLASTP (version 2.8.1) analysis of *A. thaliana* and *A. lyrata* (<https://www.arabidopsis.org>, TAIR10; <http://genome.jgi.doe.gov>, Phytozome v12: Alyrata_384_v2.1; Rawat et al., 2015; Boratyn et al., 2013). The coding sequences were aligned according to the amino acid sequences aligned by MAFFT (version 7.407) (Katoh and Standley, 2013). To evaluate the functional divergence between *A. thaliana* and *A. lyrata*, the nonsynonymous and synonymous substitutions in the 20,498 orthologs were counted. The K_A/K_S ratio was calculated according to Yang and Nielsen's method in the “yn00” program of PAML (version 4.8a) (Yang and Nielsen, 2000; Yang, 2007).

We compared the K_A/K_S ratios of mQTGs, mQTGs with CNVs, and randomly selected genes (**Figure 1B** and **Supplementary Table S1**). The mQTGs were found to have significantly higher K_A/K_S ratios than the randomly selected genes (Wilcoxon rank

sum test: $P < 0.001$; **Figure 1B**), indicating that functional divergence was more commonly detected for mQTGs than for the other genes. Additionally, the K_A/K_S ratios were higher for mQTGs with CNVs than for mQTGs and randomly selected genes (Wilcoxon rank sum test: $P < 0.001$; **Figure 1B**), suggesting that CNVs enhanced the functional divergence of mQTGs between *A. thaliana* and *A. lyrata*.

Selection Pressure for CNVs in a Species Lineage

A strong positive selection decreases the nucleotide diversity around target sites throughout the genome (i.e., selective sweep). The mQTGs with CNVs are more frequently affected by a selective sweep than the other genes in *A. thaliana* accessions (Shirai et al., 2017). This suggests that CNVs contribute to local adaptations at the intra-species level of evolution. The results of the present study suggest that CNVs contribute to the functional divergence of mQTGs at the inter-species and intra-species levels. However, it remains unclear whether positive or relaxed selection pressure controls mQTGs with CNVs at the inter-species level of evolution.

In earlier investigations, determining the selection pressure generally involved comparisons between variations at the inter-species and intra-species levels of evolution (McDonald and Kreitman, 1991; Rand and Kann, 1996; Smith and Eyre-Walker, 2002; Stoletzki and Eyre-Walker, 2011). These studies compared the number of nonsynonymous mutations (P_n), the number of synonymous mutations (P_s), the number of nonsynonymous substitutions (D_n), and the number of synonymous substitutions (D_s). The neutrality index (NI; i.e., $P_{n/s}/D_{n/s}$) is one of the parameters for comparing the variations and inferring the selection pressure (Rand and Kann, 1996). The NI quantifies the direction and extent of the difference from neutrality in which $P_{n/s}$ equals $D_{n/s}$. That is, an NI of 1 means the intra-species and inter-species functional divergences are the same. Additionally,

NI < 1 and NI > 1 reflect greater inter-species and intra-species functional divergences, respectively. Moreover, NI < 1 and NI > 1 represent the effects of positive and negative selection, respectively. We calculated the NI based on the variations of mQTGs with CNVs within *A. thaliana* accessions (intra-species) and between *A. thaliana* and *A. lyrata* (inter-species) among 20,214 genes. The P_n and P_s were estimated according to the SNPs of the 270 accessions ($dN_{\text{SNP}}/dS_{\text{SNP}}$ calculation). The D_n and D_s were estimated based on the substitutions of the orthologs between *A. thaliana* and *A. lyrata* (K_A/K_S calculation).

We found that mQTGs and mQTGs with CNVs tend to have a lower NI than the randomly selected genes (Wilcoxon rank sum test: $P < 0.05$; **Figure 1C** and **Supplementary Table S1**). These results indicate that mQTGs and mQTGs with CNVs enhanced the inter-species functional divergence over the intra-species functional divergence. To address whether mQTGs with CNVs are associated with positive selection due to functional divergence, we examined the proportion of mQTGs with CNVs in positively selected genes and in other genes. We defined positively selected genes as a gene with NI < 1 and a significant difference between $P_{n/s}$ and $D_{n/s}$ (false discovery rate < 0.05 according to the chi-squared test; **Supplementary Table S1**). The proportion of mQTGs with CNVs (0.37% = 4/1,076) was significantly higher for positively selected genes than for the other genes (0.22% = 45/19,138) (chi-squared test: $P = 2.42 \times 10^{-49}$; **Supplementary Table S2**). These results imply that CNVs tend to be contained in the mQTGs related to the adaptive evolution of *A. thaliana*.

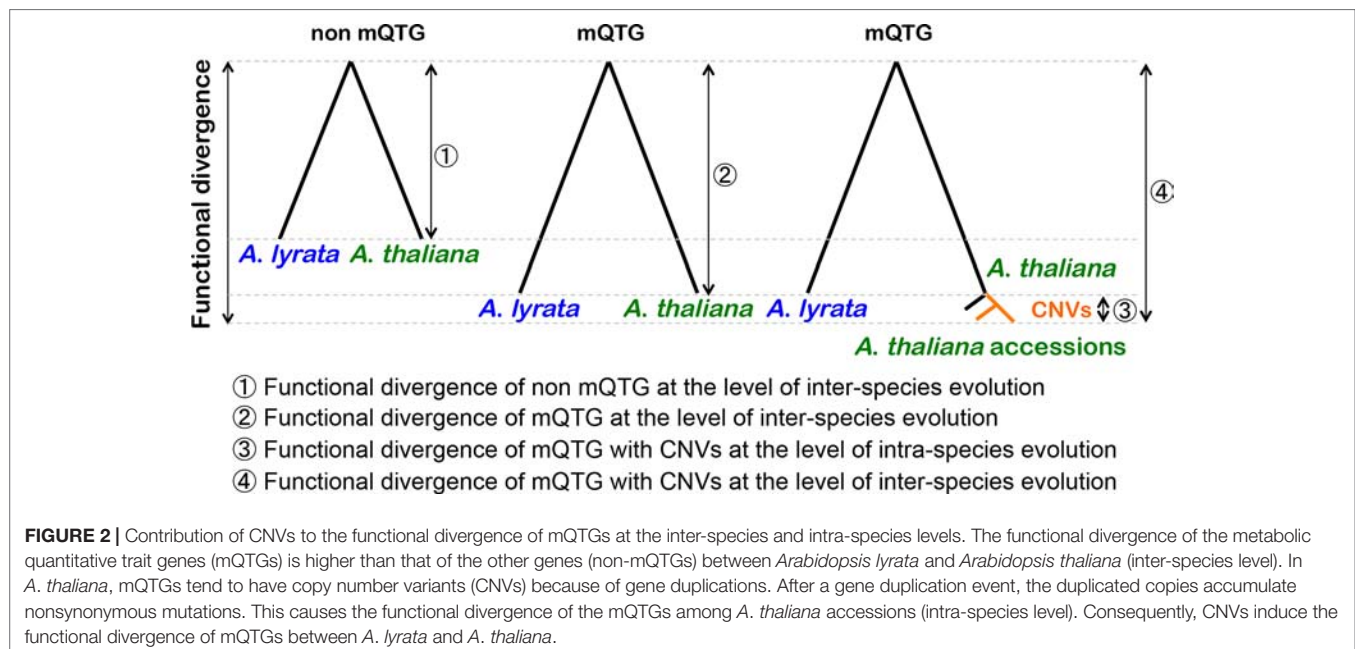
The NI reportedly leads to the incorrect determination of natural selection when there is an insufficient number of substitutions and mutations (Stoletzki and Eyre-Walker, 2011). Therefore, we validated the inferred selection pressure based on the direction of selection (DoS) (Stoletzki and Eyre-Walker, 2011). The DoS was defined as $D_n/(D_n + D_s) - P_n/(P_n + P_s)$. Additionally, DoS > 0 and DoS < 0 represent the effect of positive and negative

selection, respectively. We defined positively selected genes as genes with DoS > 0 and a significant difference between $P_{n/s}$ and $D_{n/s}$ (false discovery rate < 0.05 according to the chi-squared test; **Supplementary Table S1**). We examined the proportion of mQTGs with CNVs in positively selected genes and in other genes. Similar to the results of our analyses of NI, the proportion of mQTGs with CNVs (0.37% = 4/1,090) was significantly higher for positively selected genes than for the other genes (0.23% = 46/19,440) (chi-squared test: $P = 6.12 \times 10^{-46}$; **Supplementary Table S3**). Thus, the DoS analysis supported the NI results.

Conclusion and Perspectives

The current study examined the relationship between CNVs and the functional divergence of mQTGs at the inter-species and intra-species levels of evolution (**Figure 2**). Gene duplications induce nonsynonymous mutations *via* relaxed selection pressures. The CNVs derived from gene duplications seem to have accelerated nonsynonymous mutations. Thus, the mQTGs with CNVs have a high functional divergence at the intra-species level of evolution. Additionally, this intra-species functional divergence increases the inter-species functional divergence of the mQTGs. In fact, the functional divergence of mQTGs with CNVs tends to be high between *A. thaliana* and *A. lyrata*. Therefore, CNVs contribute to the functional divergence related to the diversity in specialized metabolites at the inter-species and intra-species levels. Consequently, CNVs tend to contribute to adaptations at the inter-species and intra-species levels. We propose that CNV is an important adaptive mechanism for generating diverse specialized metabolites in plants.

Our analyses are based on SNP calling with short-read sequencing. When SNPs are predicted in genes with CNVs based on the short reads, the SNPs are detected in the representative sequence of copied genes. The SNP detection over- or



under-estimates the number of SNPs depending on the number of copied genes. In this study, we focused on only the rate of nonsynonymous or synonymous mutations. It is unlikely that the miscalling of SNPs between nonsynonymous and synonymous mutations is biased. Therefore, we believe that the effect of miscalling is limited for our analyses.

In the past 10 years, short-read sequencing has mainly been applied in investigations at the genome scale. Unfortunately, detecting structural variants is difficult based on short-read sequencing (van Dijk et al., 2018). Therefore, there have been relatively few studies on the CNVs in plants. However, third-generation sequencing platforms, such as Pacific Bioscience (PacBio), that can generate long reads (> 5 kb) have recently been applied for plant genomic research (Zhang et al., 2016; Fukushima et al., 2017; Lan et al., 2017; Baek et al., 2018; Edger et al., 2019). The long-read sequencing data may enable the accurate detection of structural variants (Jiao and Schneeberger, 2017; van Dijk et al., 2018). For example, structural variants were recently detected by PacBio in a tropical maize inbred line (Yang et al., 2019). If this experimental approach becomes more affordable, CNVs in plants may be more easily detected. Therefore, in the near future, it will be possible to verify conclusions in other plant species.

DATA AVAILABILITY STATEMENT

The datasets for this study are available in the 1001 Genomes (<http://1001genomes.org>), TAIR10 (<https://www.arabidopsis.org>), and Phytozome v12 (<http://genome.jgi.doe.gov>) databases.

REFERENCES

- Alseikh, S., Tohge, T., Wendenberg, R., Scossa, F., Omranian, N., Li, J., et al. (2015). Identification and mode of inheritance of quantitative trait loci for secondary metabolite abundance in tomato. *Plant Cell* 27, 485–512. doi: 10.1105/tpc.114.132266
- Baek, S., Choi, K., Kim, G. B., Yu, H. J., Cho, A., Jang, H., et al. (2018). Draft genome sequence of wild *Prunus yedoensis* reveals massive inter-specific hybridization between sympatric flowering cherries. *Genome Biol.* 19, 1–17. doi: 10.1186/s13059-018-1497-y
- Bunsupa, S., Hanada, K., Maruyama, A., Aoyagi, K., Komatsu, K., Ueno, H., et al. (2016). Molecular evolution and functional characterization of a bifunctional decarboxylase involved in lycopodium alkaloid biosynthesis. *Plant Physiol.* 171, 2432–2444. doi: 10.1104/pp.16.00639
- Cao, J., Schneeberger, K., Ossowski, S., Günther, T., Bender, S., Fitz, J., et al. (2011). Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat. Genet.* 43, 956–963. doi: 10.1038/ng.911
- Carreno-Quintero, N., Bouwmeester, H. J., and Keurentjes, J. J. B. (2013). Genetic analysis of metabolome-phenotype interactions: From model to crop species. *Trends Genet.* 29, 41–50. doi: 10.1016/j.tig.2012.09.006
- Chan, E. K. F., Rowe, H. C., and Kliebenstein, D. J. (2010). Understanding the evolution of defense metabolites in *Arabidopsis thaliana* using genome-wide association mapping. *Genetics* 185, 991–1007. doi: 10.1534/genetics.109.108522
- Chye, M. L., Li, H. Y., and Yung, M. H. (2000). Single amino acid substitutions at the acyl-CoA-binding domain interrupt 14[C]palmitoyl-CoA binding of ACBP2, an *Arabidopsis* acyl-CoA-binding protein with ankyrin repeats. *Plant Mol. Biol.* 44, 711–721. doi: 10.1023/A:1026524108095
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila*

AUTHOR CONTRIBUTIONS

KS analyzed the data and wrote the manuscript. KS and KH designed the data analysis method and revised and approved the manuscript.

FUNDING

This work was supported by Grants-in-Aid for Scientific Research (25710017, 15H02433, 17H03727, 18KK0176, 18H02420, and 19H05348; to KH) as well as research grants from the Takeda Science Foundation (to KH), the Sumitomo Foundation (to KH), Kurume Research Park (to KH), and the Asahi Glass Foundation (to KH).

ACKNOWLEDGMENTS

We thank the National Institute of Genetics of the Research Organization of Information and Systems for providing excellent supercomputer services. We also thank Edanz Group (www.edanzediting.com/ac) for editing a draft of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.01567/full#supplementary-material>

- melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*. 6, 80–92. doi: 10.4161/fly.19695
- Edger, P. P., Poorten, T. J., VanBuren, R., Hardigan, M. A., Colle, M., McKain, M. R., et al. (2019). Origin and evolution of the octoploid strawberry genome. *Nat. Genet.* 51, 541–547. doi: 10.1038/s41588-019-0356-4
- Fukushima, K., Fang, X., Alvarez-Ponce, D., Cai, H., Carretero-Paulet, L., Chen, C., et al. (2017). Genome of the pitcher plant *Cephalotus* reveals genetic changes associated with carnivory. *Nat. Ecol. Evol.* 1, 1–9. doi: 10.1038/s41559-016-0059
- Gan, X., Stegle, O., Behr, J., Steffen, J. G., Drewe, P., Hildebrand, K. L., et al. (2011). Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. *Nature* 477, 419–423. doi: 10.1038/nature10414
- Hanada, K., Zou, C., Lehti-Shiu, M. D., Shinozaki, K., and Shiu, S.-H. (2008). Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. *Plant Physiol.* 148, 993–1003. doi: 10.1104/pp.108.122457
- Hanada, K., Kuromori, T., Myouga, F., Toyoda, T., and Shinozaki, K. (2009). Increased expression and protein divergence in duplicate genes is associated with morphological diversification. *PLoS Genet.* 5, 1–7. doi: 10.1371/journal.pgen.1000781
- Jiao, W. B., and Schneeberger, K. (2017). The impact of third generation genomic technologies on plant genome assembly. *Curr. Opin. Plant Biol.* 36, 64–70. doi: 10.1016/j.pbi.2017.02.002
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Kliebenstein, D. J., and Osbourn, A. (2012). Making new molecules - evolution of pathways for novel metabolites in plants. *Curr. Opin. Plant Biol.* 15, 415–423. doi: 10.1016/j.pbi.2012.05.005
- Kliebenstein, D. J. (2001). Gene duplication in the diversification of secondary metabolism: tandem 2-oxoglutarate-dependent dioxygenases control

- glucosinolate biosynthesis in *Arabidopsis*. *Plant Cell Online* 13, 681–693. doi: 10.1105/tpc.13.3.681
- Lan, T., Renner, T., Ibarra-Laclette, E., Farr, K. M., Chang, T.-H., Cervantes-Pérez, S. A., et al. (2017). Long-read sequencing uncovers the adaptive topography of a carnivorous plant genome. *Proc. Natl. Acad. Sci.* 114, E4435–E4441. doi: 10.1073/pnas.1702072114
- Lye, Z. N., and Purugganan, M. D. (2019). Copy number variation in domestication. *Trends Plant Sci.* 24, 352–365. doi: 10.1016/j.tplants.2019.01.003
- Matsuda, F., Nakabayashi, R., Yang, Z., Okazaki, Y., Yonemaru, J. I., Ebana, K., et al. (2015). Metabolome-genome-wide association study dissects genetic architecture for generating natural variation in rice secondary metabolism. *Plant J.* 81, 13–23. doi: 10.1111/tpj.12681
- McDonald, J. H., and Kreitman, M. (1991). Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351, 652–654. doi: 10.1038/351652a0
- Mouille, G., Witucka-wall, H., Bruyant, M., Loudet, O., Pelletier, S., Pauly, M., et al. (2006). Quantitative trait loci analysis of primary cell wall composition in *Arabidopsis* 1. *Plant Physiol.* 141, 1035–1044. doi: 10.1104/pp.106.079384.aestivum
- Nei, M., and Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* 3, 418–426. doi: 10.1093/oxfordjournals.molbev.a040410
- Ohno, S. (1970). *Evolution by Gene Duplication*.
- Panchy, N., Lehti-Shiu, M. D., and Shiu, S.-H. (2016). Evolution of gene duplication in plants. *Plant Physiol.* 171, 2294–2316. doi: 10.1104/pp.16.00523
- Pichersky, E., and Lewinsohn, E. (2011). Convergent evolution in plant specialized metabolism. *Annu. Rev. Plant Biol.* 62, 549–566. doi: 10.1146/annurev-arplant-042110-103814
- Rand, D., and Kann, L. (1996). Excess amino acid polymorphism in mitochondrial among genes from *Drosophila*, Mice, and Humans. *Mol. Biol. Evol.* 13, 735–748. doi: 10.1093/oxfordjournals.molbev.a025634
- Rawat, V., Abdelsamad, A., Pietzenek, B., Seymour, D. K., Koenig, D., Weigel, D., et al. (2015). Improving the annotation of *Arabidopsis* Lyrata using RNA-seq data. *PLoS One* 10, 1–12. doi: 10.1371/journal.pone.0137391
- Scannell, D. R., and Wolfe, K. H. (2008). A burst of protein sequence evolution and a prolonged period of asymmetric evolution follow gene duplication in yeast. *Genome Res.* 18, 137–147. doi: 10.1101/gr.6341207
- Schmitz, R. J., Schultz, M. D., Urich, M., Nery, J. R., Pelizzola, M., Libiger, O., et al. (2013). Patterns of population epigenomic diversity. *Nature* 495, 193–198. doi: 10.1038/nature11968
- Shirai, K., Matsuda, F., Nakabayashi, R., Okamoto, M., Tanaka, M., Fujimoto, A., et al. (2017). A highly specific genome-wide association study integrated with transcriptome data reveals the contribution of copy number variations to specialized metabolites in *Arabidopsis thaliana* accessions. *Mol. Biol. Evol.* 34, 3111–3122. doi: 10.1093/molbev/msx234
- Smith, N. G. C., and Eyre-Walker, A. (2002). Adaptive protein evolution in *Drosophila*. *Nature* 415, 1022–1024. doi: 10.1038/4151022a
- Stoletzki, N., and Eyre-Walker, A. (2011). Estimation of the neutrality index. *Mol. Biol. Evol.* 28, 63–70. doi: 10.1093/molbev/msq249
- van Dijk, E. L., Jaszczyszyn, Y., Naquin, D., and Thermes, C. (2018). The third revolution in sequencing technology. *Trends Genet.* 34, 666–681. doi: 10.1016/j.tig.2018.05.008
- Weigel, D. (2012). Natural variation in *Arabidopsis*: from molecular genetics to ecological genomics. *Plant Physiol.* 158, 2–22. doi: 10.1104/pp.111.189845
- Yang, Z., and Nielsen, R. (2000). Estimating Synonymous and Nonsynonymous Substitution Rates Under Realistic Evolutionary Models. *Mol. Biol. Evol.* 17, 32–43. doi: 10.1093/oxfordjournals.molbev.a026236
- Yang, N., Liu, J., Gao, Q., Gui, S., Chen, L., Yang, L., et al. (2019). Genome assembly of a tropical maize inbred line provides insights into structural variation and crop improvement. *Nat. Genet.* 51, 1052–1059. doi: 10.1038/s41588-019-0427-6
- Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088
- Yu, Q., Jalaludin, A., Han, H., Chen, M., Sammons, R. D., and Powles, S. B. (2015). Evolution of a double amino acid substitution in the 5-enolpyruvylshikimate-3-phosphate synthase in *Eleusine indica* conferring high-level glyphosate resistance. *Plant Physiol.* 167, 1440–1447. doi: 10.1104/pp.15.00146
- Zhang, J., Chen, L. L., Sun, S., Kudrna, D., Copetti, D., Li, W., et al. (2016). Data descriptor: building two indica rice reference genomes with PacBio long-read and illumina paired-end sequencing data. *Sci. Data* 3, 1–8. doi: 10.1038/sdata.2016.76
- Zmienko, A., Samelak, A., Kozłowski, P., and Figlerowicz, M. (2014). Copy number polymorphism in plant genomes. *Theor. Appl. Genet.* 127, 1–18. doi: 10.1007/s00122-013-2177-7

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Shirai and Hanada. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Molecular Basis of C-30 Product Regioselectivity of Legume Oxidases Involved in High-Value Triterpenoid Biosynthesis

OPEN ACCESS

Edited by:

Dae-Kyun Ro,
University of Calgary,
Canada

Reviewed by:

Yansheng Zhang,
Chinese Academy of Sciences,
China

Bjoern Hamberger,
Michigan State University,
United States

*Correspondence:

Toshiya Muranaka
muranaka@bio.eng.osaka-u.ac.jp

[†]Current address:

Hiroshi Sudo,
School of Pharmacy and
Pharmaceutical Sciences,
Hoshi University, Shinagawa, Japan
Kiyoshi Ohyama,
Leaf Tobacco Research Center,
Japan Tobacco Inc.,
Oyama, Japan

Specialty section:

This article was submitted to
Plant Metabolism and
Chemodiversity,
a section of the journal
Frontiers in Plant Science

Received: 15 June 2019

Accepted: 31 October 2019

Published: 26 November 2019

Citation:

Fanani MZ, Fukushima EO, Sawai S,
Tang J, Ishimori M, Sudo H,
Ohyama K, Seki H, Saito K and
Muranaka T (2019) Molecular Basis
of C-30 Product Regioselectivity of
Legume Oxidases Involved in High-
Value Triterpenoid Biosynthesis.
Front. Plant Sci. 10:1520.
doi: 10.3389/fpls.2019.01520

Much Zaenal Fanani¹, Ery Odette Fukushima^{1,2}, Satoru Sawai^{1,3,4,5}, Jianwei Tang³,
Masato Ishimori⁴, Hiroshi Sudo^{5†}, Kiyoshi Ohyama^{3,6†}, Hikaru Seki^{1,3}, Kazuki Saito^{3,4}
and Toshiya Muranaka^{1,3*}

¹ Department of Biotechnology, Graduate School of Engineering, Osaka University, Suita, Japan, ² Department of Biotechnology, Faculty of Life Sciences, Universidad Regional Amazónica IKIAM, Tena, Ecuador, ³ RIKEN Center for Sustainable Resource Science, Yokohama, Japan, ⁴ Graduate School of Pharmaceutical Sciences, Chiba University, Chiba, Japan, ⁵ Tokiwa Phytochemical Co., Ltd., Sakura, Japan, ⁶ Department of Chemistry and Materials Science, Tokyo Institute of Technology, Meguro, Japan

The triterpenes are structurally diverse group of specialized metabolites with important roles in plant defense and human health. Glycyrrhizin, with a carboxyl group at C-30 of its aglycone moiety, is a valuable triterpene glycoside, the production of which is restricted to legume medicinal plants belonging to the *Glycyrrhiza* species. Cytochrome P450 monooxygenases (P450s) are important for generating triterpene chemodiversity by catalyzing site-specific oxidation of the triterpene scaffold. CYP72A154 was previously identified from the glycyrrhizin-producing plant *Glycyrrhiza uralensis* as a C-30 oxidase in glycyrrhizin biosynthesis, but its regioselectivity is rather low. In contrast, CYP72A63 from *Medicago truncatula* showed superior regioselectivity in C-30 oxidation, improving the production of glycyrrhizin aglycone in engineered yeast. The underlying molecular basis of C-30 product regioselectivity is not well understood. Here, we identified two amino acid residues that control C-30 product regioselectivity and contribute to the chemodiversity of triterpenes accumulated in legumes. Amino acid sequence comparison combined with structural analysis of the protein model identified Leu149 and Leu398 as important amino acid residues for C-30 product regioselectivity. These results were further confirmed by mutagenesis of CYP72A154 homologs from glycyrrhizin-producing species, functional phylogenomics analyses, and comparison of corresponding residues of C-30 oxidase homologs in other legumes. These findings could be combined with metabolic engineering to further enhance the production of high-value triterpene compounds.

Keywords: chemodiversity, cytochrome P450 monooxygenase, legume, product regioselectivity, triterpene

INTRODUCTION

The triterpenoids are a large group of plant specialized metabolites consisting of six isoprene units. Plants produce structurally diverse triterpenoids that often have important roles in plant defense (Osbourn, 1996; Kuzina et al., 2009; Liu et al., 2019). Moreover, some triterpenoids exhibit properties beneficial for human health (Ito et al., 1988; Kenarova et al., 1990; Zhao et al., 2006;

Kojoma et al., 2010). Due to this structural diversity, triterpenoids are considered important sources for new drug leads (Geisler et al., 2013; Vo et al., 2017). However, harnessing the potential of the structural diversity of triterpenoids has been hampered by limited information regarding the molecular mechanisms underlying their structural diversity.

Decoration of the triterpene scaffold catalyzed by cytochrome P450 monooxygenases (P450s) is the second step of triterpene biosynthesis (Seki et al., 2015). Generally, P450s have the ability to stereo- and regioselectively oxidize non-activated carbon by introducing various functional groups, such as hydroxyl, carbonyl, carboxyl, and even epoxy moieties (Qi et al., 2006; Ghosh, 2017). Moreover, the introduction of a hydroxyl group into the triterpene scaffold allows to the generation of glycosylated and acylated triterpenes (Osborn et al., 2011; Seki et al., 2015). Therefore, P450s are believed to play important roles in the diversity of triterpene structures (Ghosh, 2017; Miettinen et al., 2017).

Glycyrrhizin is a triterpene saponin that is the main active compound in legume medicinal *Glycyrrhiza* plants (Shibata, 2000). In addition to its sweet taste (150 times sweeter than sucrose; Kitagawa, 2002), glycyrrhizin also shows various pharmacological activities (Shibata, 2000), including anti-inflammatory (Kroes et al., 1997), hepatoprotective (Jeong et al., 2002), and antiviral effects (Ito et al., 1988). Among the *Glycyrrhiza* species, *Glycyrrhiza uralensis*, *Glycyrrhiza glabra*, and *Glycyrrhiza inflata* are known to produce glycyrrhizin (Hayashi et al., 2000). Large amounts of glycyrrhizin accumulate in their roots and stolons, accounting for an estimated 2%–8% of the dry weight (Shibata, 2000). Glycyrrhizin itself has been used as an ingredient in a number of commercial products, including foods, personal health care products, and medicines. However, the production of glycyrrhizin is dependent on natural resources that require an approximately 2–3-year growth period before harvesting (Chen et al., 2014). Due to the economic value and market demand for licorice, overexploitation of wild licorice has led to significant environmental issues (Marui et al., 2011). Therefore, a rapid and environmental friendly system for glycyrrhizin production is required.

Metabolic engineering has been studied extensively for production of plant specialized metabolites in engineered organisms. The biosynthesis of glycyrrhizin involves the initial cyclization of 2,3-oxidosqualene to the pentacyclic triterpene β -amyrin, followed by a series of oxidative reactions at positions C-11 and C-30 (Seki et al., 2008; Seki et al., 2011). Previously, we identified two P450s (CYP88D6 and CYP72A154) involved in glycyrrhizin biosynthesis (Seki et al., 2008; Seki et al., 2011). Functional characterization of CYP72A154 showed that this enzyme catalyzed oxidation at C-30, accompanied by the production of isomers as minor products (Seki et al., 2011). *Medicago truncatula* does not produce glycyrrhizin, and the homologous enzyme, CYP72A63, showed superior regioselective oxidation of the C-30 position only (Seki et al., 2011). Use of CYP72A63 for production of glycyrrhizin aglycone in engineered yeast further enhanced its yield (Zhu et al., 2018). Although improvement of glycyrrhizin aglycone

production in yeast has been achieved, it still produces a number of byproducts, such as 11 α ,30-dihydroxy- β -amyrin and 30-hydroxy- β -amyrin (Zhu et al., 2018). Further improvement of glycyrrhizin production by combining metabolic engineering and protein engineering is hampered by limited knowledge regarding the structure–function relationship of P450s involved in glycyrrhizin biosynthesis.

A great deal of research effort has focused on discovering the P450s involved in decoration of the triterpene skeleton. Based on amino acid sequence identity, CYP72A154 and CYP72A63 are classified as members of the CYP72A subfamily. The CYP72A subfamily is known as P450 subfamily enzyme involved in generating triterpene chemodiversity, where they catalyze site-specific oxidation of the oleanane-type triterpenoid scaffolds, C-2 β (Biazzi et al., 2015), C-21 β (Yano et al., 2017; Leveau et al., 2019), C-22 β (Ebizuka et al., 2011; Fukushima et al., 2013), C-23 (Fukushima et al., 2013; Biazzi et al., 2015; Liu et al., 2019), and C-30 (Seki et al., 2011). Prall et al. (2016) reported that the CYP72A subfamily in flowering plants showed high variability of amino acid residues among substrate recognition sites (SRSs). However, there have been no experimental studies investigating the roles of amino acid residues in the SRSs in this subfamily. Moreover, no CYP72A subfamily protein crystal structures have yet been reported, even for closely related P450 with more than 40% identity. Therefore, it is still difficult to apply rational protein engineering to the CYP72A subfamily to improve product specificity and regioselectivity. Some reports suggested that gene mining of publicly available genomic or transcriptome databases is a more practical method for obtaining candidate genes encoding CYP72A enzymes showing better catalytic activity (Moses et al., 2014; Suzuki et al., 2018; Zhu et al., 2018). However, the majority of natural enzymes still exhibit some properties unfavorable for application in metabolic engineering, such as low product specificity and regioselectivity (Jung et al., 2011; Goldsmith and Tawfik, 2017). Therefore, it would be useful and interesting to determine the molecular mechanism underlying product regioselectivity of C-30 oxidases involved in high-value triterpenoid biosynthesis.

In this study, we mined the CYP72A subfamily from *M. truncatula* and characterized its enzymatic activity against β -amyrin. Functional characterization of the CYP72A subfamily from *M. truncatula* showed that CYP72A63 is an enzyme with high C-30 product regioselectivity. Interestingly, CYP72A62v2 and CYP72A64v2, which share more than 90% amino acid sequence identity, showed completely different product regioselectivity. By comparing the SRS sequences of CYP72A63 and its homologs, and by protein homology modeling, we identified Leu149 and Leu398 as key amino acid residues responsible for C-30 regioselective oxidation activity in CYP72A63. Analysis of CYP72A154 variants from both glycyrrhizin-producing and non-glycyrrhizin-producing *Glycyrrhiza* species also indicated that amino acid residue #398 differentiated the product regioselectivity of CYP72A154 variants. The results of this study will provide opportunities to engineer P450s for manipulation of product regioselectivity by rational protein engineering, to achieve the production of valuable triterpenoids such as glycyrrhizin.

MATERIALS AND METHODS

Plant Materials

The seed plants used in this experiment are listed in **Supplementary Table 1**. Seeds were germinated by mechanical scarification and imbibition in the dark at 23°C for 2 days. Germinated seeds were planted in soil and grown in a plant room with a controlled temperature of 23°C under a 16-h light/8-h dark photoperiod. Plant samples for RNA isolation were collected from 4-week-old seedlings, immediately frozen in liquid nitrogen, and then stored at -80°C until use. Underground parts of *Glycyrrhiza pallidiflora* were harvested from the Medicinal Plant Garden of Chiba University (Chiba, Japan); *Glycyrrhiza lepidota* and *Glycyrrhiza macedonica* were obtained from Osaka University of Pharmaceutical Science (Osaka, Japan), and *Glycyrrhiza glabra* was obtained from Health Sciences University of Hokkaido (Hokkaido, Japan).

Authentic Standards

β-Amyrin was purchased from Extrasynthese (Lyon, France). Sophoradiol, 30-hydroxy-β-amyrin, and 11-deoxoglycyrrhetinic acid were synthesized as described in our previous report (Seki et al., 2008).

Gene Mining for the CYP72A Subfamily

CYP72A subfamily candidates were identified by BLAST search using the amino acid sequences of CYP72A61v2, CYP72A63, and CYP72A67 as queries against the *M. truncatula* genome project Mt4.0v1 proteins (Tang et al., 2014). Hits showing >50% identity with unique sequential gene ID numbers were manually checked for surrounding 50 kb in genome JBrowse (Krishnakumar et al., 2015). Natural variants of gene cluster subgroup III were mined from *M. truncatula* Hapmap (Zhou et al., 2017) using the genomic sequences of CYP72A62 (Medtr8g042060), CYP72A63 (Medtr8g042040), CYP72A64 (Medtr8g042020), and CYP72A65 (Medtr8g042000) (obtained from *M. truncatula* genome database) as queries. Amino acid sequences were predicted according to the known coding sequences of gene references in *M. truncatula*. CYP72A63 homologs from other legumes were mined from publicly available genomic sequences in the 1KP database (Matasci et al., 2014), Clover GARDEN (www.clovergarden.jp/) (Hirakawa et al., 2016), Vigna Genome Server (viggs.dna.affrc.go.jp/) (Sakai et al., 2016), Cool Season Food Legume Genome Database (www.coolseasonfoodlegume.org/) (Main et al., 2013), and Legume Information System (legumeinfo.org/) (Dash et al., 2016).

Cloning and Vector Construction

RNA preparation and cloning methods are explained briefly in the **Supplementary Methods**. All candidates were verified by sequencing and submitted to the P450 Committee for naming. Yeast expression clones, using pELC-CPR-GW (Seki et al., 2008), pYES2-DEST52 (Thermo Fisher Scientific, Waltham, MA), and a Gateway-compatible version of pESC-HIS (Seki et al., unpublished) as destination vectors, were

constructed by LR reaction using LR clonase II Enzyme mix (Thermo Fisher Scientific).

Accession Numbers

Sequence data from this experiment have been submitted to the DNA DataBank of Japan (DDBJ), the European Nucleotide Archive (ENA), and GenBank databases under the following accession numbers: MK534530 (CYP72A695), MK534531 (CYP72A696), MK534532 (*Gmax*CYP72A141), MK534533 (*Gg*CYP72A154), MK534534 (*Gp*CYP72A154), MK534535 (*Lc*CYP72A698), MK534536 (*Ps*CYP72A698), MK534537 (CYP72A302), MK534538 (CYP72A694), MK534539 (CYP72A697), MK534540 (CYP72A336v2), MK534541 (CYP72A70), MK534542 (*Gsoja*CYP72A141), MK534543 (CYP72A337v2), MK534544 (CYP72A557), MK534545 (CYP72A558), MK534546 (CYP72A559), MK534547 (CYP72A560), MK534548 (CYP72A64v2), MK534549 (CYP72A699), MK792941 (CYP72A66v2).

In Vivo Enzymatic Assay in Yeast

In vivo enzymatic assay in yeast was performed by co-expression of *Lotus japonicus* cytochrome P450 reductase (CPR) and β-amyrin synthase (OSC1), with three yeast expression clones for each of the CYP72A subfamily genes. Each corresponding set of pELC-CPR-CYP72A, pYES2-DEST52-CYP72A, and pESC-HIS-CYP72A was transformed into *Saccharomyces cerevisiae* INVSc1 (*MATa his3D1 leu2 trp1-289 ura3-52*; Thermo Fisher Scientific) carrying pYES3-ADH-OSC1 (Seki et al., 2008) using Frozen-EZ Yeast Transformation II™ (Zymo Research, Orange, CA), and a yeast strain harboring the three empty vectors (pELC-CPR-GW, pYES2-DEST52, and pESC-HIS) was used as a control. All genes tested in this experiment are listed in **Supplementary Table 3**. *In vivo* enzymatic assay in yeast was performed as reported previously (Fukushima et al., 2011) with some modifications. Yeast strains were pre-cultured in appropriate synthetic defined medium (Clontech, Palo Alto, CA) containing 2% glucose, and incubated overnight at 30°C, 200 rpm. Aliquots of 50 μl of yeast pre-cultures were added into 5 ml of appropriate synthetic defined medium (Clontech) containing 2% glucose and incubated overnight at 30°C, 200 rpm. Yeast cells were collected by centrifugation, resuspended in 5 ml of appropriate synthetic defined medium (Clontech) containing 2% galactose and incubated at 30°C for 4 days at 200 rpm. Yeast metabolites were extracted using ethyl acetate after sonication, three times for 30 minutes each time, and portions of the extracts were analyzed by gas chromatography-mass spectrometry (GC-MS) after derivatization with *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide (Sigma-Aldrich, St. Louis, MO). GC-MS analysis was performed using a gas chromatograph (7890B; Agilent Technologies, Santa Clara, CA) connected to a mass spectrometer (5977A; Agilent Technologies) and HP-5MS capillary column (0.25 mm × 30 m, 0.25 μm) (Agilent Technologies). The initial oven temperature was 150°C with a hold time of 1 min, increasing from 150°C to 260°C at 30°C/min and 260°C to 300°C at 1°C/min. Samples were injected in splitless mode with an injection temperature of 250°C, with helium as the carrier gas at a flow rate of 1.0 ml/min. Comparison of retention

times and mass fragmentation patterns of detectable compounds with those of authentic standards was performed to assign the peaks.

Bioinformatics Analyses

The predicted amino acid sequences of identified CYP72A subfamily members were used for multiple sequence alignment using ClustalW in MEGA7 (Kumar et al., 2016). A neighbor-joining tree was generated using MEGA6 with the Jones–Taylor–Thomson substitution model and bootstrap analysis of 1,000 replicates. The SRSs of the *M. truncatula* CYP72A subfamily were predicted as described previously (Gotoh, 1992; Prall et al., 2016). A CYP72A63 structure model was constructed using 6C93.A as the template with about 27% identity in SWISS-MODEL (Arnold et al., 2006). Heme was docked with 40 × 40 × 40 grid points, spacing 0.375, and grid center -14.738; 21.827; 9.827 using Autodock4.0 (Morris et al., 2009). PyMol (Schrödinger, 2017) was used to visualize the amino acids in the active site of the enzyme.

Site-directed Mutagenesis

Designed protein variants were generated using the site-specific nucleotides listed in **Supplementary Table 2**. Mutagenesis experiments were performed using a PrimeSTAR Mutagenesis Basal Kit (TaKaRa Bio, Kyoto, Japan) and the entry clone was used as a template.

Elucidation of the Structure of Compound 2 (Peak 2)

A yeast strain carrying β -amyrin synthase co-expressed together with CPR and CYP72A63^{L149V/L398V} was cultured in appropriate synthetic defined medium with a total volume of 25.2 L (250 ml × 72, 150 ml × 48). Yeast metabolites were obtained by saponification prior to extraction with *n*-hexane three times. Yeast extracts were evaporated and the residues were applied to silica gel chromatography (60N, spherical, neutral) (Kanto Chemical, Tokyo, Japan). Hexane:ethyl acetate (1:9) was used as the mobile phase, and afforded about 8 mg of compound 2. Nuclear magnetic resonance (NMR) data were recorded on a Bruker Avance III 600 MHz spectrometer (Bruker Daltonics, Bremen, Germany) using CDCl₃ as the solvent.

RESULTS

The Chromosomal Localization of *M. truncatula* CYP72As Corresponds to the Phylogenetic Tree Topology

Using representatives of each subgroup in the CYP72A subfamily described previously (Seki et al., 2011), we identified 20 candidate genes encoding CYP72A subfamily enzymes in the *M. truncatula* genome. CYP72A59-like6, CYP72A59-like7, and CYP72A68-like showed shorter sequences and no transcripts in transcriptome data (**Supplementary Table 4**). We considered

them to be pseudogenes and they were excluded from subsequent experiments. Among the 17 genes identified as encoding CYP72A subfamily enzymes from *M. truncatula* Mt4.0v1, only eight (CYP72A59v2, CYP72A61, CYP72A62v2, CYP72A63, CYP72A65v2, CYP72A67, CYP72A68-430, CYP72A68-470) were functionally characterized (Seki et al., 2011; Fukushima et al., 2013; Biazzi et al., 2015; Reed et al., 2017). In summary, nine genes (CYP72A64v2, CYP72A66v2, CYP72A336v2, CYP72A337v2, CYP72A70, CYP72A557, CYP72A558, CYP72A559, CYP72A560) were reported in this study.

In the BlastP results, hits had unique sequential gene ID numbers (**Supplementary Table 5**). Based on this observation, we next analyzed the chromosomal localization of the CYP72A subfamily in the *Medicago truncatula* Mt4.01 genome database using JBrowse (Krishnakumar et al., 2015). The CYP72A subfamily genes were shown to be clustered in tandem arrays on chromosomes 2 and 8, with the exception of CYP72A67, CYP72A61, and CYP72A70 (**Figure 1A**). Notably, the gene cluster on chromosome 2 contains half of the total number of CYP72A subfamily genes present in the *M. truncatula* genome. The CYP72A subfamily enzymes within the cluster showed 81%–94% identity except for CYP72A337 (**Supplementary Table 6**). The constructed phylogenetic tree showed that the CYP72A subfamily genes clustered on chromosomes 2 and 8 are also grouped into the same subgroups (IV and II, respectively; **Figure 1B**). Consistent with its amino acid sequence identity, CYP72A337v2 is located out of the clade within the gene cluster, and was further classified into a new subgroup IV. These results suggested that gene duplication likely occurred multiple times within the CYP72A subfamily in *M. truncatula*.

CYP72A63 Regioselectively Oxidized β -Amyrin C-30

The enzymatic activities of CYP72A subfamily enzymes against β -amyrin were characterized by co-expression of CYP72A subfamily genes together with CPR and β -amyrin synthase in *S. cerevisiae* INVSc1. The CYP72A subfamily enzymes within the cluster showed a range of regioselective oxidation activity (**Figure 2**, **Supplementary Figure 1A**). CYP72A61v2 of subgroup I showed oxidation activity against β -amyrin at position C-22 β , producing sophoradiol. β -Amyrin C-22 β oxidation activity was also detected in CYP72A66v2 (subgroup III), which showed oxidation activity at more than one site producing sophoradiol, 30-hydroxy- β -amyrin, compound 2 (peak 2 in **Figure 2**, which may correspond to a monohydroxylated β -amyrin product with a hydroxyl group on the D or E ring, based on the characteristics of retro Diels–Alder fragmentation at the C ring in the mass fragmentation pattern shown in **Supplementary Figure 1A**), 11-deoxoglycyrrhetinic acid, and some minor unknown compounds. CYP72A557 and CYP72A558 also showed oxidation activity against β -amyrin producing unknown compound 1 (peak 1 in **Figure 2**), compound 2 (peak 2 in **Figure 2**), and unknown compound 3 (peak 3 in **Figure 2**), which were predicted to be monohydroxylated β -amyrin products with a hydroxyl group on the D or E, based on the characteristics of retro Diels–Alder

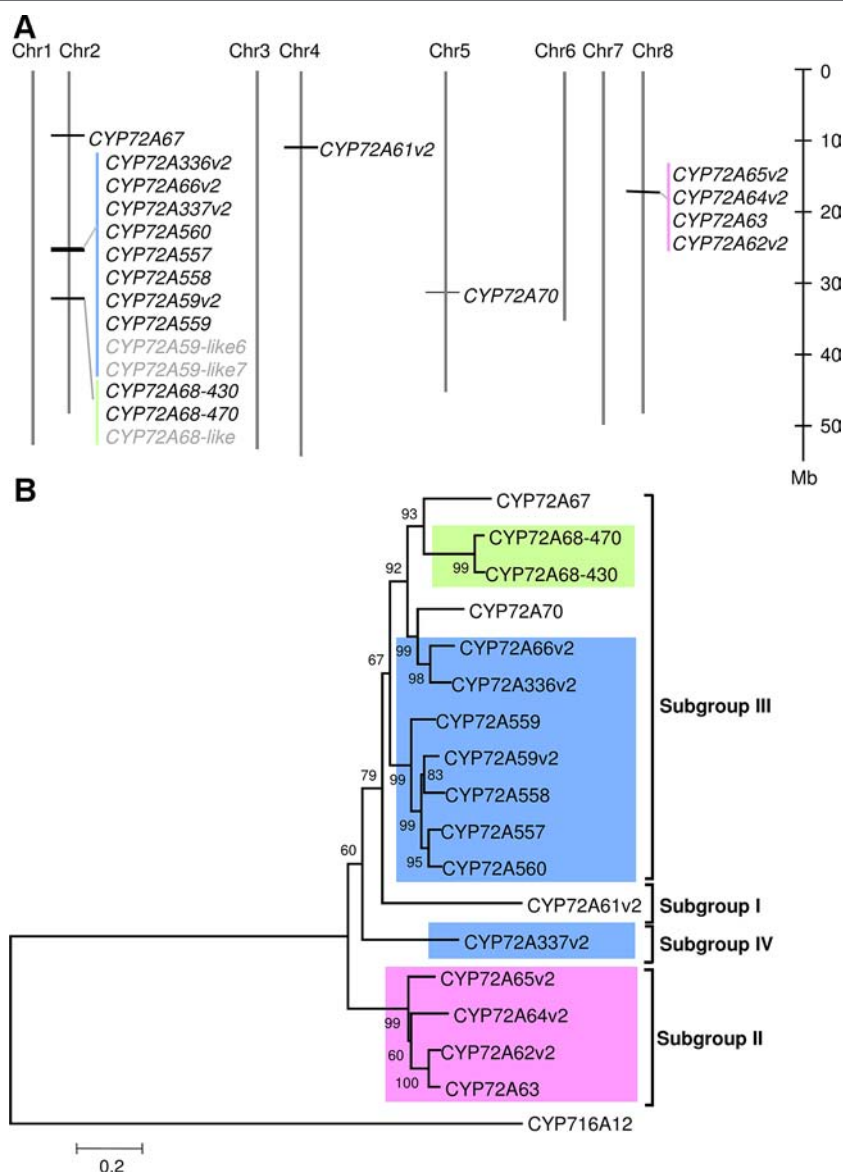


FIGURE 1 | CYP72A subfamily genes in *M. truncatula*. **(A)** Chromosomal localization of CYP72A subfamily in *M. truncatula*. Chromosomal localization was determined according to the Mtr4.0 database JBrowser feature. Each cluster is indicated in a different color. P450s shown in gray indicate pseudogenes. **(B)** Phylogenetic tree of CYP72A subfamily. A neighbor-joining tree was constructed using amino acid sequences with bootstrap analysis of 1,000 replicates. Members of the CYP72A subfamily located in the same cluster are indicated by the same background color. The subgroup classification based on Seki et al. (2011) is indicated.

fragmentation at the C ring in the mass fragmentation pattern shown in **Supplementary Figure 1A**. Unlike other enzymes in the same cluster, CYP72A559 and CYP72A560 showed oxidation activity against β -amyrin at the D or E ring, to produce unknown compound **1** (peak 1 in **Figure 2**) alone.

CYP72A subfamily enzymes in tandem array subgroup II showed oxidation activity with different regioselectivities (**Figure 2**, **Supplementary Figure 1A**). CYP72A65v2 showed oxidation activity against β -amyrin, producing unknown compound **1** (peak 1 in **Figure 2**) as the major product and compound **2** (peak 2 in **Figure 2**) as a minor product. CYP72A62v2 and CYP72A64v2 showed oxidation activity against β -amyrin, producing compound

2 (peak 2 in **Figure 2**) and unknown compound **4** (peak 4 in **Figure 2**) as minor carboxylated form products. In addition, unknown compound **1** (peak 1 in **Figure 2**) was detected in trace amounts in the reaction product of CYP72A62v2. CYP72A63 showed oxidation activity against β -amyrin at the C-30 position producing 30-hydroxy- β -amyrin and 11-deoxoglycyrrhetic acid. The *in vivo* enzymatic assay clearly showed that CYP72A63 was the only one CYP72A subfamily enzyme with high regioselectivity at β -amyrin C-30 from *M. truncatula*.

No oxidized β -amyrin product was detected on *in vivo* enzyme assay of CYP72A70, CYP72A336v2, CYP72A59v2, and CYP72A337v2 (**Figure 2**, **Supplementary Figure 1A**). The lack

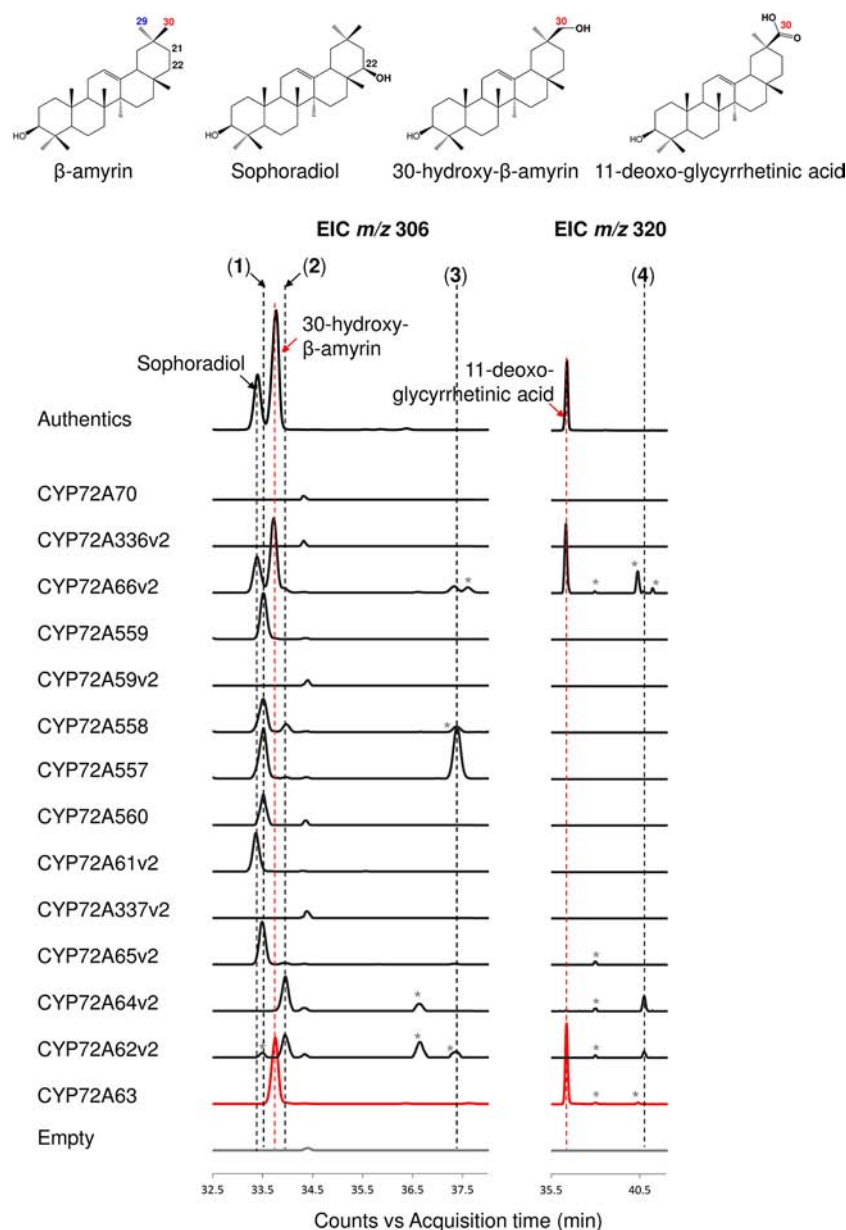


FIGURE 2 | *In vivo* enzymatic activity of CYP72A subfamily enzymes from *M. truncatula* against β -amyrin as the substrate. Triterpene extract was analyzed by gas chromatography coupled with mass spectrometry (extracted ion chromatogram; EIC). Molecular ions with m/z 306 and 320 were selected as specific for hydroxylated and carboxylated β -amyrin, respectively. Peaks corresponding to authentic standards are labeled. Peaks likely corresponding to β -amyrin oxidized products are labeled 1–4. Peaks corresponding to unconfirmed β -amyrin oxidized products are indicated with asterisks. Red color indicates C-30 product regioselectivity-related product or enzymes.

of detectable enzymatic activity of these four CYP72A enzymes may have different causes; current detection method may not be able to detect trace amounts of its oxidation products, enzymes may not be expressed in correct way, enzymes may have different substrate specificity or mutation in the signature region may cause loss-of-function. Possible mutations in the signature region of these four enzymes were investigated by multiple alignment of the putative oxygen activation region of the CYP72A subfamily (**Supplementary Figure 3**). An amino

acid substitution in a conserved acidic amino acid (Glu) to basic amino acid (Lys) was found in CYP72A336v2. To examine whether this substitution of conserved amino acid residue Glu327 caused loss of function in CYP72A336v2, the mutant CYP72A336v2^{K327E} was generated. *In vivo* enzymatic assay of CYP72A336v2^{K327E} showed that substitution of Lys327 to Glu327 could recover enzyme activity of CYP72A336v2, producing sophoradiol, 30-hydroxy- β -amyrin, and compound 2 (peak 2 in **Supplementary Figure 3C**).

Leu149 and Leu398 are Essential for Regioselective Oxidation at C-30

To identify residues important for C-30 regioselectivity of CYP72A63, we examined amino acid residues in the predicted SRS (Figure 3A). A number of criteria were applied; the amino acid residue must be conserved in CYP72A62v2 and CYP72A64v2, but not in CYP72A63, or must not be conserved at all. Six positions (Val149, His150, Ile244, Gln246, Glu269, and Val398) were selected (marked in the box, Figure 3A). *In vivo* enzymatic assay of CYP72A62v2 mutants showed that CYP72A62v2^{V398L} was sufficient to alter the product regioselectivity by detecting compound 2 (peak 2) and 30-hydroxy- β -amyrin on GC-MS

analysis (Figure 3C, Supplementary Figure 1B). In parallel, we also mapped the positions of key residues important for C-30 product regioselectivity by generating protein chimeras of CYP72A63 and CYP72A62v2 using the segment exchange approach (Supplementary Figure 4). *In vivo* enzymatic assay of chimeric proteins suggested that key amino acid residues for C-30 product regioselectivity are located between residues #133 and #409, and more than one amino acid residue is required (Supplementary Figure 4). These results suggested that Leu398 is not the sole residue responsible for C-30 product regioselectivity.

To identify the second important amino acid residue determining product regioselectivity, we next mapped the

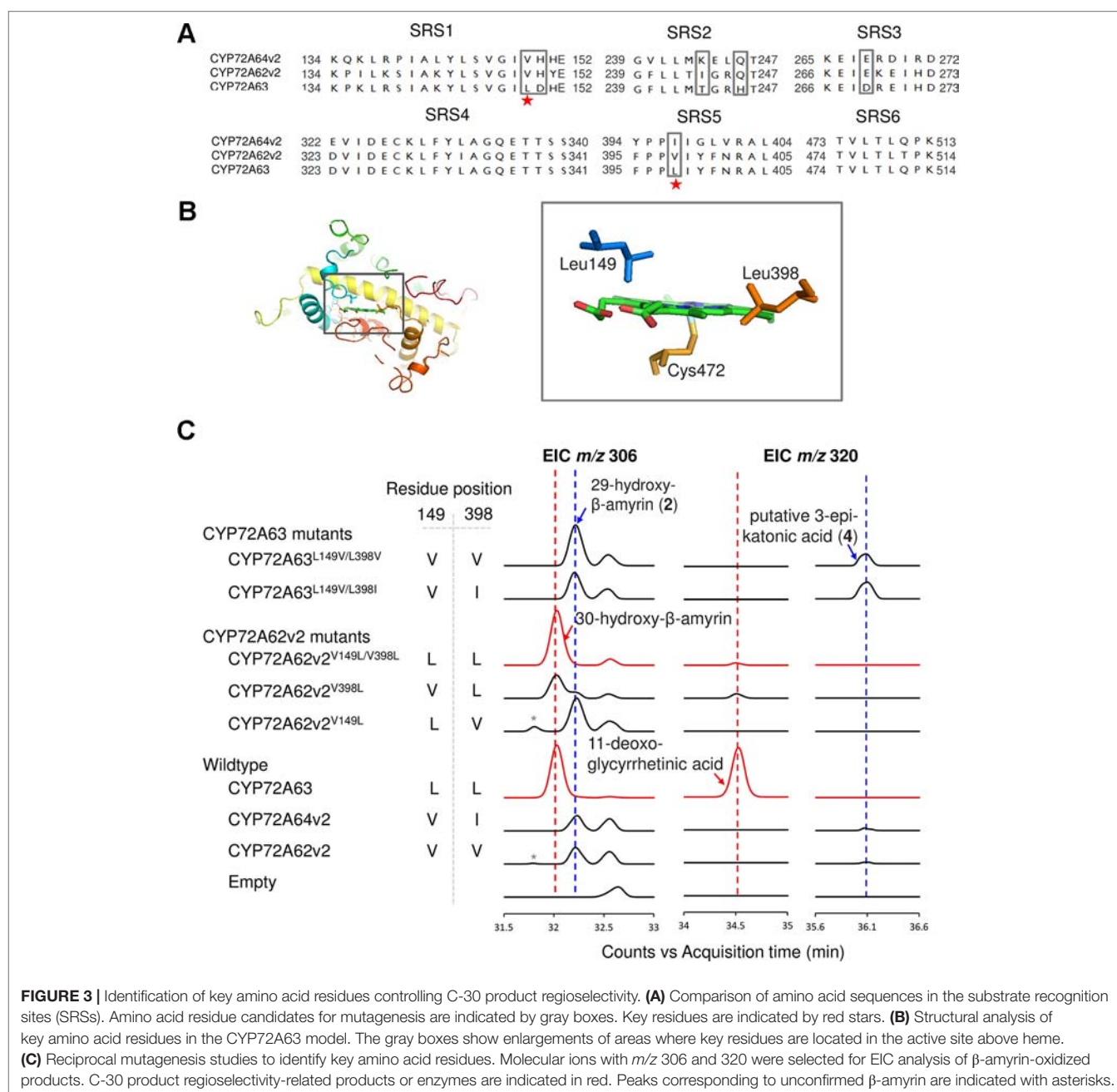


FIGURE 3 | Identification of key amino acid residues controlling C-30 product regioselectivity. **(A)** Comparison of amino acid sequences in the substrate recognition sites (SRSs). Amino acid residue candidates for mutagenesis are indicated by gray boxes. Key residues are indicated by red stars. **(B)** Structural analysis of key amino acid residues in the CYP72A63 model. The gray boxes show enlargements of areas where key residues are located in the active site above heme. **(C)** Reciprocal mutagenesis studies to identify key amino acid residues. Molecular ions with m/z 306 and 320 were selected for EIC analysis of β -amyrin-oxidized products. C-30 product regioselectivity-related products or enzymes are indicated in red. Peaks corresponding to unconfirmed β -amyrin are indicated with asterisks.

position of Leu398 in the three-dimensional homology model of CYP72A63 (**Figure 3B**). The homology model of CYP72A63 showed that Leu398 is located in the area surrounding the reaction center of P450, where the enzyme catalytic reaction takes place. Based on these findings, we hypothesized that the second important amino acid residue may be located close to Leu398. Therefore, we examined amino acid residues surrounding Leu398. Among five amino acid residues in this region, Leu149 (SRS1) is located in a relatively face-to-face position to Leu398 (SRS5) (**Figure 3B**). Thus, both amino acid residues may determine the regioselectivity of CYP72A63. To examine the role of Leu149 together with Leu398 in C-30 product regioselectivity, the mutant CYP72A62v2^{V149L/V398L} was generated. *In vivo* enzymatic assay of CYP72A62v2^{V149L/V398L} showed that CYP72A62v2^{V149L/V398L} produced only C-30 oxidized products, 30-hydroxy- β -amyrin and 11-deoxoglycyrrhetinic acid (**Figure 3C**). These results showed that Leu149 and Leu398 are important amino acid residues for regioselective oxidation of β -amyrin at C-30.

Substitution of amino acid residues #149 and #398 markedly altered the product regioselectivity of the CYP72A62v2 enzyme (**Figure 3C**, **Supplementary Figure 1C**). We also generated mutants of CYP72A63, designated as CYP72A63^{L149V/L398V} and CYP72A63^{L149V/L398I}, which resembled CYP72A62v2 and CYP72A64v2, respectively. *In vivo* enzymatic assay showed that CYP72A63^{L149V/L398V} and CYP72A63^{L149V/L398I} had altered product regioselectivity resembling their counterparts, CYP72A62v2 and CYP72A64v2, respectively (**Figure 3C**, **Supplementary Figure 1C**). These results clearly showed that Leu149 and Leu398 are important amino acid residues for C-30 product regioselectivity of CYP72A63. To determine the regioselective oxidation activities of CYP72A62v2 and CYP72A64v2, we examined the structure of compound 2 (peak 2) by NMR spectroscopy. Complete ¹³C assignment of purified compound 2 (peak 2) was not achieved due to incomplete removal of impurities. However, the data indicated the presence of 29-hydroxy- β -amyrin (**Supplementary Data**).

We used the *M. truncatula* Hapmap to further investigate the roles of residues #149 and #398 in product regioselectivity (Zhou et al., 2017). By focusing on amino acid residues #149 and #398, amino acid residue variants present in CYP72A62v2 were identified from *M. truncatula* accessions (Zhou et al., 2017) (**Figure 4A**). The variants were classified into three types based on differences in amino acid residues #149 and #398: Type VV (Val149, Val398), Type VL (Val149, Leu398), and Type IL (Ile149, Leu398). To evaluate the effects of divergent amino acid residues on product regioselectivity, we generated mutants of CYP72A63 mimicking each Types, VV, VL, and IL. *In vivo* enzymatic assay showed that these artificial mutant enzymes had differences in product regioselectivity (**Figure 4B**, **Supplementary Figure 1D**). Unexpectedly, Type IL oxidized at the C-30 position, producing 30-hydroxy- β -amyrin and 11-deoxoglycyrrhetinic acid, which resembled CYP72A63 rather than CYP72A62v2. These results clearly showed that amino acid residues #149 and #398 are essential for fine tuning of product regioselectivity.

Residue #398 may be Involved in Generating Triterpene Chemodiversity in *Glycyrrhiza* Species

To investigate the roles of amino acid residues #149 and #398 in generating triterpene chemodiversity in *Glycyrrhiza* species, CYP72A154 variants from glycyrrhizin-producing species (*Glycyrrhiza uralensis*, GuCYP72A154; *G. glabra*, GgCYP72A154) and non-glycyrrhizin-producing species (*Glycyrrhiza pallidiflora*, GpCYP72A154; *G. lepidota*, GlCYP72A154; *G. macedonica*, GmacCYP72A154) were investigated. As the full-length amino acid sequences of the three non-glycyrrhizin-producing species were identical, we selected GpCYP72A154 as a representative species. Sequence alignment of CYP72A154 variants showed that amino acid residue #149 (numbering based on CYP72A63) is Val149 for both types, but amino acid residue #398 (numbering based on CYP72A63) differs between them, i.e., Gly398 for glycyrrhizin-producing species and Ala398 for non-glycyrrhizin-producing species (**Figure 5**). To characterize the role of divergent amino acid residue #398 in product regioselectivity in CYP72A154 variants, *in vivo* enzymatic assay was performed. GuCYP72A154 and GgCYP72A154 oxidized β -amyrin with less regioselectivity at the D or E ring, producing unknown compound 1 (peak 1) and 30-hydroxy- β -amyrin as the main products and 29-hydroxy- β -amyrin as a trace product (**Figure 5**, **Supplementary Figure 1E**). In contrast, the non-glycyrrhizin-producing species GpCYP72A154 oxidized β -amyrin, producing unknown compound 1 (peak 1) and 29-hydroxy- β -amyrin as the main products and 30-hydroxy- β -amyrin as a trace product (**Figure 5**, **Supplementary Figure 1E**). These results suggested that the product regioselectivity of CYP72A154 variants differs between glycyrrhizin-producing species and non-glycyrrhizin-producing species.

To characterize the product regioselectivity of CYP72A154 variants, we also elucidated two unknown products in the residue remaining after purification of 30-hydroxy-11-oxo- β -amyrin from yeast co-expressing β -amyrin synthase, CYP88D6, GuCYP72A154, and CPR (Peaks 4b and 4c in **Figure 2A** shown in Seki et al., 2011). NMR spectroscopy showed that the unknown compounds of 30-hydroxy-11-oxo- β -amyrin isomers were 29-hydroxy-11-oxo- β -amyrin and 21 β -hydroxy-11-oxo- β -amyrin (**Supplementary Data**). These results showed that GuCYP72A154 catalyzed oxidation at C-21 β , C-29, and C-30.

To investigate whether the differences in product regioselectivity of CYP72A154 variants are associated with divergence of amino acid residue #398, we generated mutant GuCYP72A154^{G398A} resembling GpCYP72A154. *In vivo* enzymatic assay showed that product regioselectivity of GuCYP72A154^{G398A} changed markedly, more closely resembling GpCYP72A154 product regioselectivity and producing unknown compound 1 (peak 1, putative 21 β -hydroxy- β -amyrin) and 29-hydroxy- β -amyrin as main products (**Figure 5**, **Supplementary Figure 1E**). In addition, we also generated mutants of CYP72A63, CYP72A63^{L149V/L398G} and CYP72A63^{L149V/L398A}, carrying the amino acids at residues #149 and #398 in glycyrrhizin-producing species GuCYP72A154 and non-glycyrrhizin-producing species GpCYP72A154 (**Figure 5**, **Supplementary Figure 1E**). *In vivo* enzymatic assay of

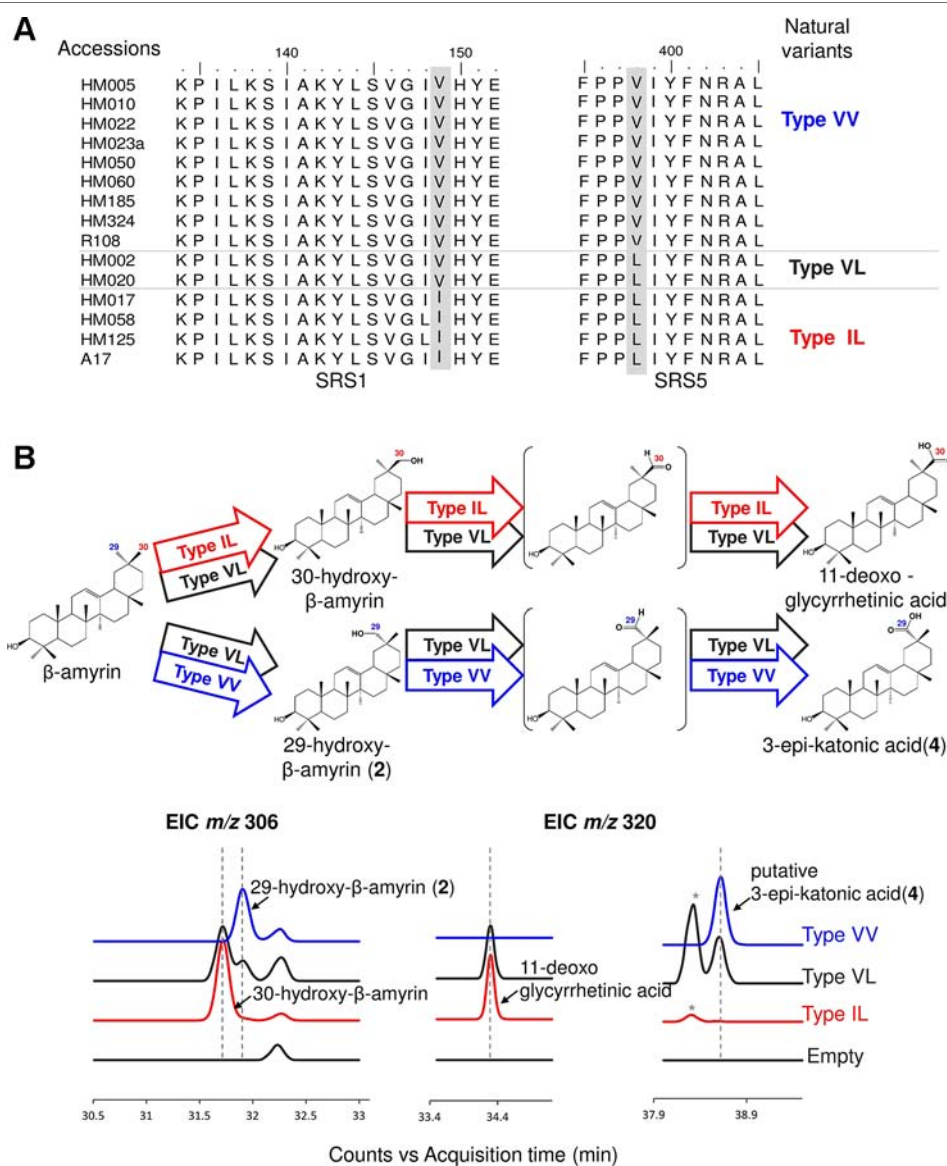


FIGURE 4 | Divergent amino acid residues #149 and #398 showed different product regioselectivity. **(A)** Amino acid sequence alignment of CYP72A62v2 natural variants obtained by gene mining. Amino acids are shown in different colors to distinguish each variant. **(B)** *In vivo* enzymatic assay of CYP72A63 mutants mimicking CYP72A62 natural variants. Different colors indicate natural variant types. Identified products are labeled. The enzymatic activity of each variant is indicated by arrows with different colors. Peaks corresponding to unconfirmed β-amyrin are indicated with asterisks.

CYP72A63^{L149V/L398G} (mimicking glycyrrhizin-producing variants) showed oxidation activity mainly at the C-30 position, producing 30-hydroxy-β-amyrin as the main product, while CYP72A63^{L149V/L398A} (mimicking non-glycyrrhizin-producing variants) showed oxidation activity mainly at the C-29 position producing 29-hydroxy-β-amyrin as the main product. Mutagenesis of CYP72A63 mimicking CYP72A154 variants showed good agreement with the product regioselectivity of CYP72A154 variants from glycyrrhizin-producing and non-producing species. These results suggested that differences in amino acid residue #398 may be involved in generating triterpene chemodiversity in *Glycyrrhiza* species by conferring variable product regioselectivity.

Divergent Amino Acid Residues #149 and #398 in Legume CYP72A63 Homologs

To further analyze the roles of amino acid residues #149 and #398 in CYP72A63 homologs from other legumes, we performed phylogenomic analyses of CYP72A63 homologs by constructing a phylogenetic tree, comparing amino acid residues #149 and #398, and performing *in vivo* enzymatic assays. CYP72A63 homologs were searched from publicly available genomic information and transcriptome databases of legume plants. Selected CYP72A63 homologs were cloned, confirmed by DNA sequencing, and submitted to the P450 Committee for naming (Supplementary Table 3). Phylogenetic analysis of CYP72A63 homologs showed

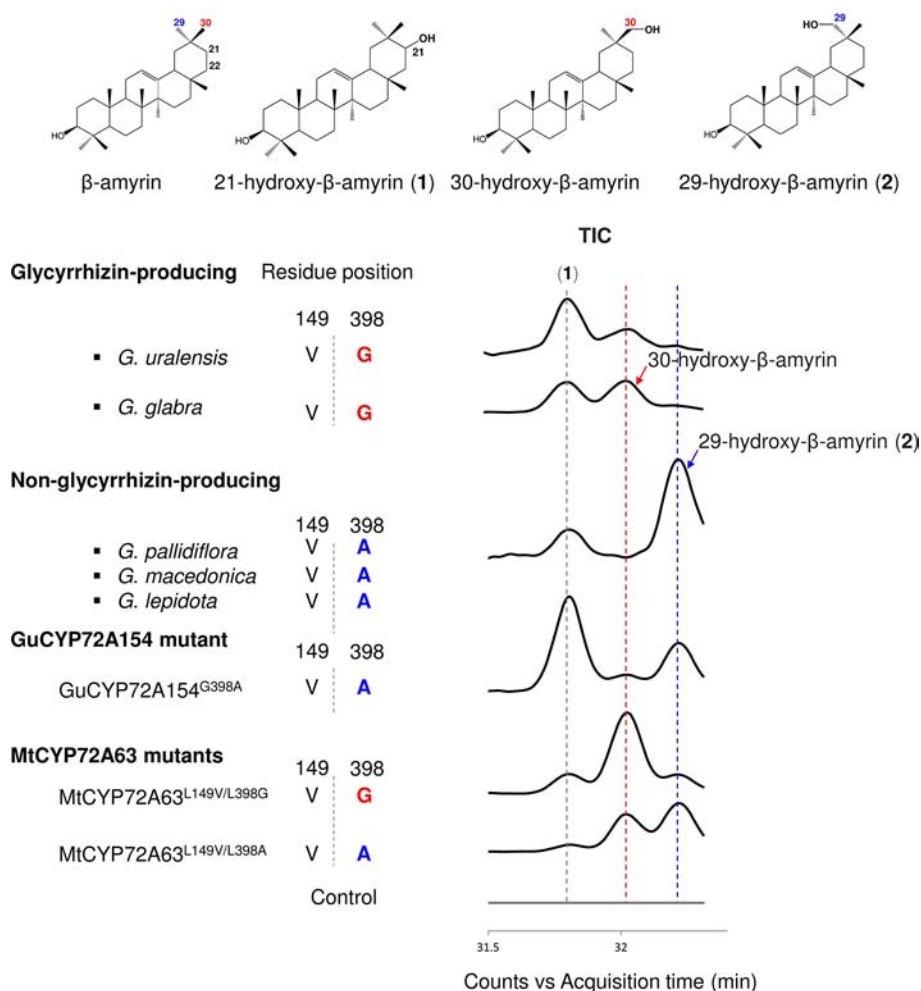


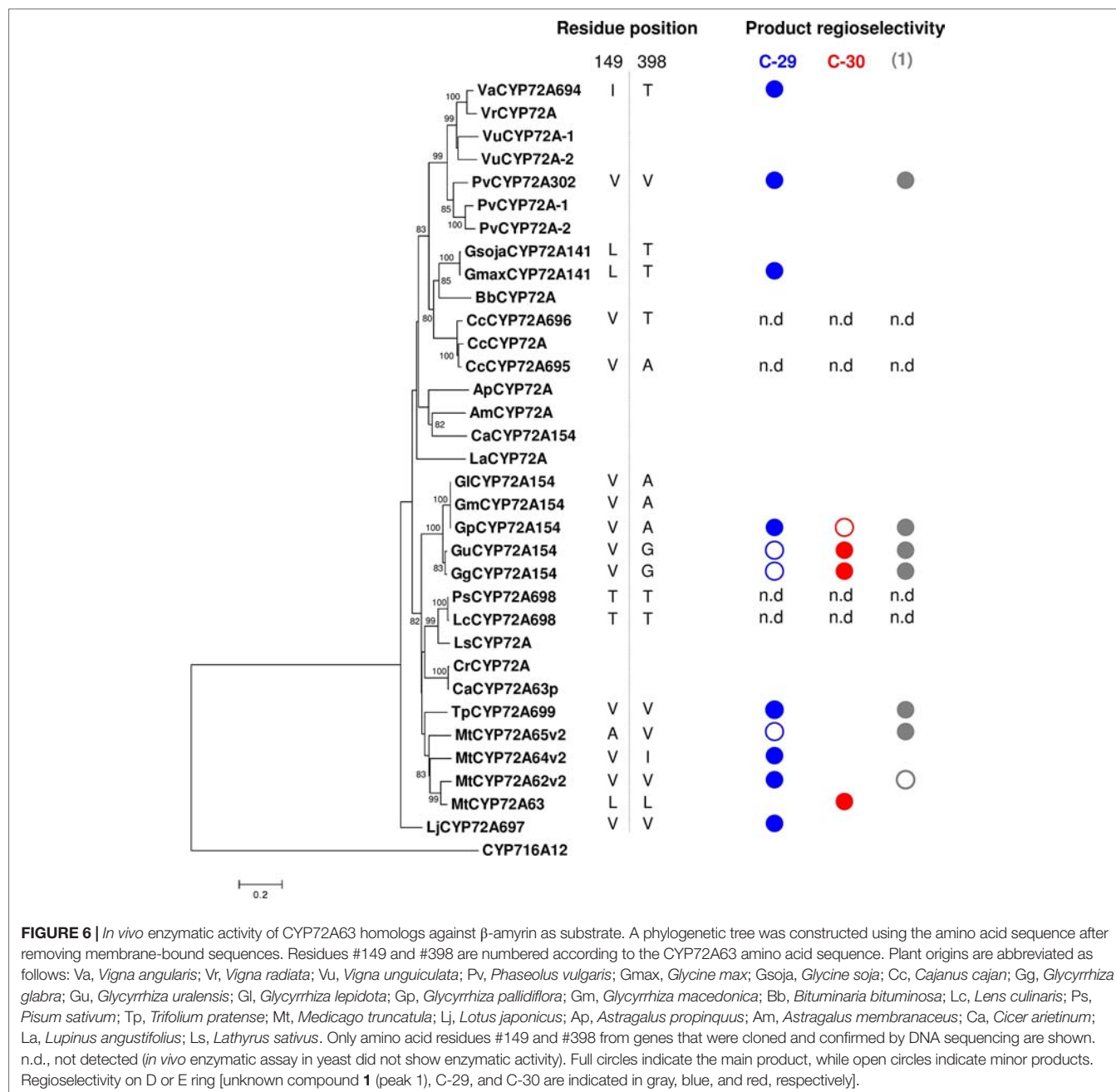
FIGURE 5 | *In vivo* enzymatic assay of CYP72A154 variants and mutants from *Glycyrrhiza* species. Amino acid residues #149 and #398 are numbered based on the CYP72A63 amino acid sequence. Total ion current chromatograms are shown in enlargement mode. Identified peaks are indicated. Amino acid residue #398 and product regioselectivity are indicated in red and blue, for C-30 and C-29, respectively.

that the legume species have a variable number of CYP72A63 homologs present in their genome (Figure 6); amino acid residues #149 and #398 varied among them. To investigate the relationships of amino acid residues #149 and #398 to product regioselectivity, *in vivo* enzymatic assays against β -amyrin were performed (Figure 6, Supplementary Figure 1F). None of the CYP72A63 homologs from these legumes showed C-30 oxidation activity, except CYP72A66v2 from *M. truncatula* and CYP72A154 from glycyrrhizin-producing *Glycyrrhiza* plants (Figure 7). The combinations of amino acid residues #149 and #398 differed among *Vigna angularis* VaCYP72A694 (Ile149, Thr398), *Glycine max* Gmax CYP72A141 (Leu149, Thr398), and *Lotus japonicus* LjCYP72A697 (Val149, Val398), but *in vivo* enzymatic assays showed that they have regioselectivity in the C-29 position. *Trifolium pratense* TpCYP72A699 and *Phaseolus vulgaris* PvCYP72A302 have a combination of amino acid residues, Val149 and Val398, as seen in *Lotus japonicus* LjCYP72A697, but *in vivo* enzymatic assay showed that they differed in product regioselectivity; however, regioselectivity

in the C-29 position was common among them. These results support the important roles of amino acid residues #149 and #398 in determining regioselective oxidation activity.

DISCUSSION

In the absence of protein structures for P450s involved in triterpene biosynthesis, comprehensive analyses of the structure–function relationships could not be performed. Recent studies successfully demonstrated that comparative functional analyses of natural variants or evolutionarily related enzymes could be useful in narrowing down important amino acid residues, such as those involved in substrate and product specificity (Komori et al., 2013; Chen and Li, 2017; Xue et al., 2018). Elucidation of the molecular basis of product specificity and regioselectivity of the enzymes enables us to improve product specificity and produce desired compounds more efficiently. In this study, investigation of tandem duplicated CYP72A subfamily genes in



M. truncatula identified two amino acid residues, Leu149 and Leu398, responsible for C-30 regioselective oxidation activity.

CYP72A subfamily genes were found in tandem arrays in the genome of *M. truncatula* (Figure 1A). The tandem array of CYP72A subfamily members is not a specific feature of *M. truncatula* because it was also found in other legumes (Supplementary Figure 2), and even in non-legumes, such as *Barbarea vulgaris* (Liu et al., 2019), *Oryza sativa*, and *Arabidopsis thaliana* (Saika et al., 2014). Recently, Liu et al. (2019) reported that CYP72A552, one of the CYP72A subfamily genes present in tandem arrays, is involved in hederagenin-based saponin

biosynthesis in *Barbarea vulgaris*. Similarly, Saika et al. (2014) also reported that CYP72A31, one of the CYP72A subfamily genes present in tandem arrays, is involved in the mechanism of herbicide tolerance in rice. These observations indicated that genes present in tandem arrays may have different functions.

Gene duplication followed by subsequent mutations is a well-known mechanism by which genes gain new functions through neofunctionalization and escape from adaptive conflict (Panchy et al., 2016). In neofunctionalization, one copy of a duplicated gene maintains the original function, while the other copy gains a novel function by accumulation

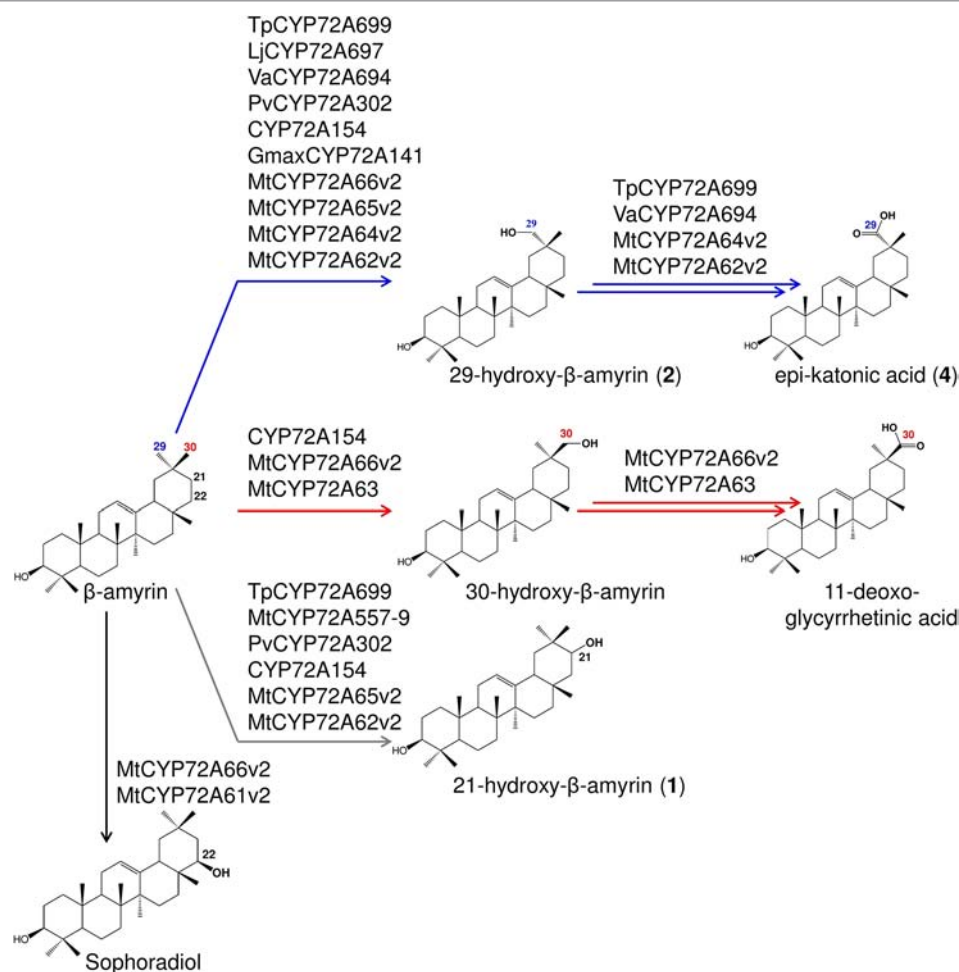


FIGURE 7 | Summary of the regioselective oxidation activity of CYP72A subfamily enzymes reported in this study. Oxidation reactions on C-22 β , putative C-21, C-29, and C-30 positions catalyzed by each P450s are indicated by arrows with different colors.

of mutations (Panchy et al., 2016). Although the reasons why members of the CYP72A subfamily are commonly found in tandem arrays are still unknown, these findings suggest a mechanism of functional diversification of CYP72A subfamily enzymes in plants. The existence of the CYP72A subfamily is not specific to legumes. However, its contribution to the synthesis of structurally diverse triterpenes has been reported almost exclusively in legumes, with the exception of the C-21 oxidase of *Avena* spp. (Leveau et al., 2019) and C-23 oxidase of *Kalopanax septemlobus* (Han et al., 2018) and *Barbarea vulgaris* (Liu et al., 2019).

CYP72A63 is located in the tandem array together with CYP72A65v2, CYP72A64v2, and CYP72A62v2 on chromosome 8. A previous report suggested that CYP72A65 also showed C-30 oxidation activity (Zhu et al., 2018). However, our *in vivo* enzymatic assay of CYP72A subfamily enzymes from *M. truncatula* clearly showed that CYP72A63 is the one enzyme that selectively oxidizes the C-30 position. Considering the chromosomal localization and phylogenetic tree topology, this tandem array likely evolved from a common

ancestor with accumulation of mutations. Mutations can directly affect enzyme function, or can be silent. In the case of CYP72A64v2 and CYP72A62v2, substitution of Ile398 to Val398 did not alter the enzyme regioselectivity. However, substitution of Val149 to Leu149 and Ile/Val398 to Leu398 changed the product regioselectivity from C-29 toward the C-30 position. Thus, amino acid residues #149 and #398 determined the substrate orientation, which controlled its product regioselectivity.

The enzymes catalyzing β -amyrin at the C-29 position were identified for the first time in this study (Figure 7). Among the C-29 oxidases, VaCYP72A694 exhibited greater accumulation of the carboxylated product (putative C-29 carboxylated product; up to 70% product ratio; Supplementary Figure 1F). The C-29-derived saponins have been found in legume plants; adzukisaponins in *V. angularis* (Kitagawa et al., 1983) and macedonocides in *G. macedonica* (Hayashi et al., 2000), and some showed promise as high-value triterpenoids. Yoshikawa et al. (2002) identified albiziasaponin B, a triterpene saponin with a carboxyl group at C-29 of its aglycone moiety, from the Thai

medicinal plant, *Albizia myriophylla* (Cha-em Thai), showing sweetness 600 times greater than sucrose. In Thai folk medicine, the stem of *A. myriophylla* has been used as a substitute for licorice due to its sweetness (Yoshikawa et al., 2002), and is one of the ingredients in traditional medicine used for treatment of diabetes (Neamsuvan et al., 2015). Thus, identification of C-29 oxidases provides a new genetic tool for production of high-value C-29-derived triterpenoids by synthetic biology.

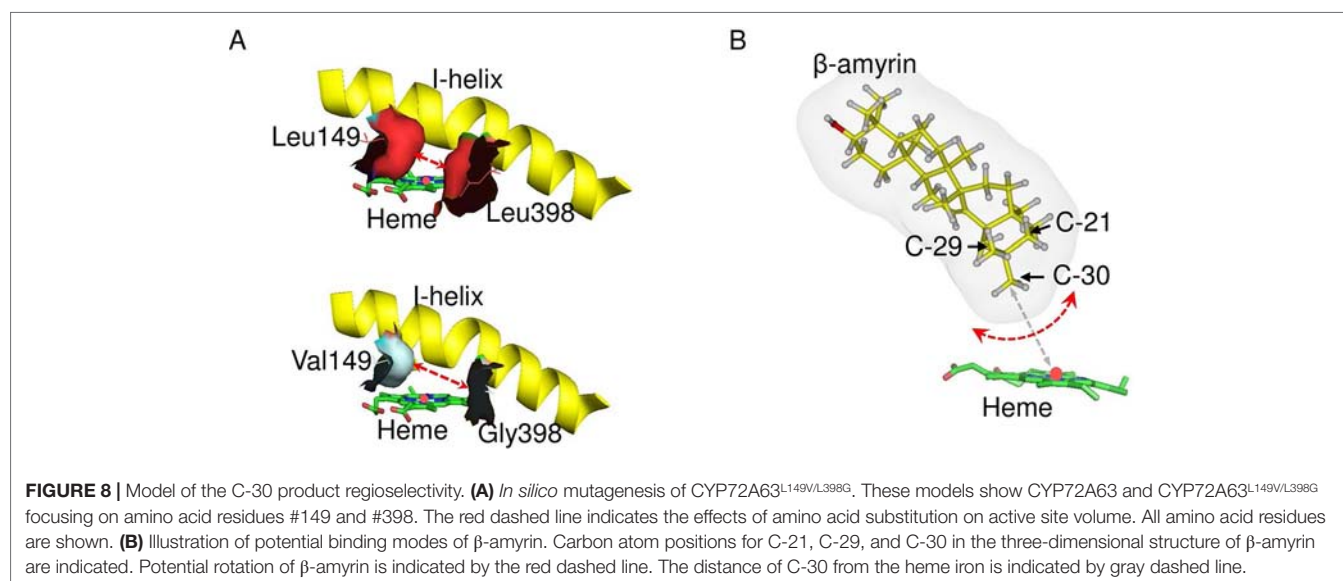
Glycyrrhiza species were classified into two types according to the accumulation of glycyrrhizin, i.e., glycyrrhizin-producing and non-glycyrrhizin-producing species. Glycyrrhizin-producing species mainly show accumulation of C-30-derived saponins, while non-producing species accumulate C-29-derived saponins (Hayashi et al., 2000). The enzymatic activity of CYP72A154 variants showed good agreement with saponin accumulation in both types of *Glycyrrhiza* species. Comparison of amino acid residues #149 and #398 in CYP72A154 variants suggested that divergence in amino acid residue #398 may be involved in the generation triterpene chemodiversity in *Glycyrrhiza* species.

A substrate–enzyme complex model could not be obtained due to the low quality of the protein model. However, C-30 product regioselectivity was illustrated by *in silico* mutagenesis of CYP72A63 (Figure 8). Regioselectivity on C-30 and C-29 positions was controlled by amino acid residues #149 and #398, but the combinations of amino acid residues were different among them. The combination of amino acid residues with long nonpolar side chains (Ile/Leu149 and Leu398) resulted in C-30 product regioselectivity, while the combination of an amino acid residue with a nonpolar short side chain (Val149) and long nonpolar side chain (Ile149) resulted in C-29 product regioselectivity. Other combinations of amino acid residues with nonpolar short side chains (Val, Gly, Ala) at both positions resulted in broad regioselectivity, producing a number of isomers, i.e., C-30, C-29, and 21 β . This suggested that C-30 product regioselectivity required amino acid residues with a nonpolar long side chain (Leu/Ile149 and Leu398) for specific

placement of the methyl-30 group in the favorable position for enzyme reaction (Figure 8). Shortening the side chains of amino acid residues #149 and #398 increased the volume of the active site cavity, which allowed positioning of the substrate in multiple orientations (Figure 8A). Thus, C-30 product isomers could be produced by nonspecific positioning of methyl-29, methyl-30, and methylene-21 functional groups in proximity to mononuclear iron (Figure 8B). Interestingly, amino acid residues #149 and #398 were nonpolar amino acids (Ile, Leu, Val, Ala), except for Gly and Thr. This suggested that hydrophobic interactions may be important for positioning of the substrate in the most favorable orientation for enzyme catalysis. To explain the effects of 20 possible amino acid residues and the binding mode of the enzyme and substrate, mutagenesis studies and crystal structure analysis of CYP72A63 are required.

A recent study showed that CYP72A69 from soybean, *G. max*, catalyzed oxidation at the C-21 β position in soyasapogenol A biosynthesis (Yano et al., 2017). *G. max* C-21 oxidase has a short side chain amino acid at residues #149 and #398 (i.e., Ala149 and Gly398, respectively; numbering based on CYP72A63). Similarly, *M. truncatula* CYP72A65v2, which catalyzed oxidation at putative C-21 (peak 1), also has a short side chain amino acid at both residues #149 and #398 (i.e., Ala149 and Val398, respectively). The substitutions Leu149Ala and Leu398Val in CYP72A63 did not alter the product regioselectivity resembling CYP72A65v2 (data not shown), suggesting that C-21 β product regioselectivity may be controlled by amino acid residues other than residues #149 and #398. The amino acid residues involved in product regioselectivity of CYP72A65v2 remain to be determined; doing so would lead to a better understanding of the evolutionary diversification of product regioselectivity of tandem array on chromosome 8.

Although we successfully altered the product regioselectivity of CYP72A62v2 from the C-29 position toward the C-30 position, by substituting Val149 to Leu149 and Val398 to Leu398, the mutant CYP72A62v2^{V149L/V398L} produced only trace amounts



of C-30 carboxylated product (**Figure 3B**). However, the original enzyme, CYP72A63, produced the carboxylated product, suggesting that additional amino acid residues, other than #149 and #398, may be required for successive oxidation to produce the C-30 carboxylated product. Komori et al. (2013) showed that an amino acid residue located in the loop of SRS6, Ser479, in CYP71V1 is important for successive oxidation of amorpha-4,11-diene to produce carboxylated product (artemisinic acid) in the biosynthesis of the antimalarial sesquiterpenoid, artemisinin, in *Artemisia annua*. This suggested that the amino acid residues involved in C-30 carboxylated product specificity may also be located in SRS regions. Further investigations are required to identify the amino acid residues involved in C-30 successive oxidation to produce the C-30 carboxylated product.

We identified the key amino acid residues, Leu149 and Leu398, controlling C-30 product regioselectivity in CYP72A63. The results reported here will enable us to improve the product specificity and produce desired compounds more efficiently. Rational engineering of C-30 oxidase by site saturation mutagenesis amino acid residues #149 and #398 may be useful for fine tuning of the methyl-30 functional group to a favorable position for enzyme catalysis, to improve product specificity and production yield. Alternatively, the results presented here suggested that it may be possible to redirect the product regioselectivity of VaCYP72A694, a high carboxylated product producer (up to 70% accumulation), toward the C-30 position by rational protein engineering. The application of protein engineering in combination with metabolic engineering has been shown to significantly improve the production of natural products. Our findings will provide opportunities to further enhance the production of the valuable triterpene glycyrrhizin through rational protein engineering of C-30 oxidase.

DATA AVAILABILITY STATEMENT

The nucleotide sequences isolated in this study have been submitted to the GenBank at NCBI, ncbi.nlm.nih.gov/genbank/.

REFERENCES

- Arnold, K., Bordoli, B., Kopp, J., and Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22, 2. doi: 10.1093/bioinformatics/bti770
- Biazzi, E., Carelli, M., Tava, A., Abbruscato, P., Losini, I., Avato, P., et al. (2015). CYP72A67 catalyzes a key oxidative step in *Medicago truncatula* hemolytic saponin biosynthesis. *Mol. Plant* 8, 10. doi: 10.1016/j.molp.2015.06.003
- Chen, H., and Li, X. (2017). Identification of a residue responsible for UDP-sugar donor selectivity of a dihydroxybenzoic acid glycosyltransferase from *Arabidopsis* natural accessions. *Plant J.* 89, 2. doi: 10.1111/tpj.13271
- Chen, K. Z., Hu, S., and Chen, R. (2014). *Licorice industry in China: implications for licorice producers in Uzbekistan*. China: International Food Policy Research Institute Beijing.
- Dash, S., Campbell, J. D., Cannon, E. K. S., Cleary, A. M., Huang, W., Kalberer, S. R., et al. (2016). Legume information system (LegumeInfo.org): a key component of a set of federated data resources for the legume family. *Nucleic Acids Res.* 44, D1. doi: 10.1093/nar/gkv1159
- Ebizuka, Y., Shibuya, M., and Wakita, E. (2011). "C-22 Hydroxylase" in *US patent 20110171698*.

AUTHOR CONTRIBUTIONS

MF, EF, and SS designed experiments. MF, SS, JT, MI, HSu, KO, and HSe performed experiments. MF, EF, SS, KO, and HSe wrote the article. EF, KS, and TM supervised the research. All authors discussed the results and approved the article.

FUNDING

This study was supported in part by the Grants-in-Aid for Scientific Research of the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number JP19H02921; the Scientific Technique Research Promotion Program for Agriculture, Forestry, Fisheries, and Food Industry, Japan; The Program for Promotion of Basic and Applied Researches for Innovations in Bio-oriented Industry (BRAIN); The Special Fund from the Director of RIKEN Yokohama Institute; The RIKEN Rijicho Fund; and the Monbukagakusho Scholarship.

ACKNOWLEDGMENTS

We thank Toshio Aoki (Nihon University) for his valuable discussions and technical advice, Mareshige Kojoma (Health Sciences University of Hokkaido, Japan) for providing *G. glabra*, Makio Shibano (Osaka University of Pharmaceutical Sciences, Japan) for providing *G. lepidota* and *G. macedonica*, David R. Nelson (University of Tennessee, USA) for the naming of P450s, Kyoko Inoue (Osaka University, Japan) for technical assistance with NMR analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <https://www.frontiersin.org/articles/10.3389/fpls.2019.01520/full#supplementary-material>

- Fukushima, E. O., Seki, H., Ohyama, K., Ono, E., Umamoto, N., Mizutani, M., et al. (2011). CYP716A subfamily members are multifunctional oxidases in triterpenoid biosynthesis. *Plant Cell Physiol.* 52, 12. doi: 10.1093/pcp/pcr146
- Fukushima, E. O., Seki, H., Sawai, S., Suzuki, M., Ohyama, K., Saito, K., et al. (2013). Combinatorial biosynthesis of legume natural and rare triterpenoids in engineered yeast. *Plant Cell Physiol.* 54, 5. doi: 10.1093/pcp/pct015
- Geisler, K., Hughes, R. K., Sainsbury, F., Lomonosoff, G. P., Rejzek, M., Fairhurst, S., et al. (2013). Biochemical analysis of a multifunctional cytochrome P450 (CYP51) enzyme required for synthesis of antimicrobial triterpenes in plants. *Proc. Natl. Acad. Sci. U.S.A.* 110, 35. doi: 10.1073/pnas.1309157110
- Ghosh, S. (2017). Triterpene structural diversification by plant cytochrome P450 enzymes. *Front. Plant Sci.* 8, 1886. doi: 10.3389/fpls.2017.01886
- Goldsmith, M., and Tawfik, D. S. (2017). Enzyme engineering: reaching the maximal catalytic efficiency peak. *Curr. Opin. Struct. Biol.* 47, 140–150. doi: 10.1016/j.sbi.2017.09.002
- Gotoh, O. (1992). Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *J. Biol. Chem.* 267, 83–90.

- Han, J. Y., Chun, J. H., Oh, S. A., Park, S. B., Hwang, H. S., Lee, H., et al. (2018). Transcriptomic analysis of *Kalopanax septemlobus* and characterization of KsBAS, CYP716A94 and CYP72A397 genes involved in hederagenin saponin biosynthesis. *Plant Cell Physiol.* 59, 319–330. doi: 10.1093/pcp/pcx188
- Hayashi, H., Hosono, N., Kondo, M., Hiraoka, N., Ikeshiro, Y., Shibano, M., et al. (2000). Phylogenetic relationship of six *Glycyrrhiza* species based on rbcL sequences and chemical constituents. *Biol. Pharm. Bull.* 23, 5. doi: 10.1248/bpb.23.602
- Hirakawa, H., Kaur, P., Shirasawa, K., Nichols, P., Nagano, S., Appels, R., et al. (2016). Draft genome sequence of subterranean clover, a reference for genus *Trifolium*. *Sci. Rep.* 6, 30358. doi: 10.1038/srep30358
- Ito, M., Sato, A., Hirabayashi, K., Tanabe, F., Shigeta, S., Baba, M., et al. (1988). Mechanism of inhibitory effect of glycyrrhizin on replication of human immunodeficiency virus (HIV). *Antiviral Res.* 10, 6. doi: 10.1016/0166-3542(88)90047-2
- Jeong, H. G., You, H. J., Park, S. J., Moon, A. R., Chung, Y. C., Kang, S. K., et al. (2002). Hepatoprotective effects of 18beta-glycyrrhetic acid on carbon tetrachloride-induced liver injury: inhibition of cytochrome P450 2E1 expression. *Pharmacol. Res.* 46, 3. doi: 10.1016/S1043-6618(02)00121-4
- Jung, S. T., Lauchli, R., and Arnold, F. H. (2011). Cytochrome P450: taming a wild type enzyme. *Curr. Opin. Biotechnol.* 22, 6. doi: 10.1016/j.copbio.2011.02.008
- Kenarova, B., Neychev, H., Hadjiivanova, C., and Petkov, V. D. (1990). Immunomodulating activity of ginsenoside Rg1 from *Panax ginseng*. *Jpn. J. Pharmacol.* 54, 4. doi: 10.1254/jip.54.447
- Kitagawa, I., Wang, H. K., Saito, M., and Yoshikawa, M. (1983). Saponin and sapogenol. XXXI. Chemical constituents of the seeds of *Vigna angularis* (WILLD.) Ohwi et Ohashi. (1). Triterpenoidal sapogenols and 3-furanmethanol β -D-glucopyranoside. *Chem. Pharm. Bull.* 31, 2. doi: 10.1248/cpb.31.664
- Kitagawa, I. (2002). Licorice root. A natural sweetener and an important ingredient in Chinese medicine. *Pure Appl. Chem.* 74, 1189–1198. doi: 10.1351/pac200274071189
- Kojoma, M., Ohyama, K., Seki, H., Hiraoka, Y., Asazu, S. N., Sawa, S., et al. (2010). *In vitro* proliferation and triterpenoid characteristics of licorice (*Glycyrrhiza uralensis* Fischer, Leguminosae) stolons. *Plant Biotechnol.* 27, 1. doi: 10.5511/plantbiotechnology
- Komori, A., Suzuki, M., Seki, H., Nishizawa, T., Meyer, J. J. M., Shimizu, H., et al. (2013). Comparative functional analysis of CYP71AV1 natural variants reveals an important residue for the successive oxidation of amorpho-4,11-diene. *FEBS Letters* 587, 3. doi: 10.1016/j.febslet.2012.11.031
- Krishnakumar, V., Kim, M., Rosen, B. D., Karamycheva, S., Bidwell, S. L., Tang, H., et al. (2015). MTGD: the *M. truncatula* genome database. *Plant Cell Physiol.* 56, 1. doi: 10.1093/pcp/pcu179
- Kroes, B. H., Beukelman, C. J., van den Berg, A. J., Wolbink, G. J., van Dijk, H., and Labadie, R. P. (1997). Inhibition of human complement by β -glycyrrhetic acid. *Immunology* 90, 1. doi: 10.1046/j.1365-2567.1997.00131.x
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 7. doi: 10.1093/molbev/msw054
- Kuzina, V., Ekström, C. T., Andersen, S. B., Nielsen, J. K., Olsen, C. E., and Bak, S. (2009). Identification of defense compounds in *Barbarea vulgaris* against the herbivore *Phyllotreta nemorum* by an ecometabolomic approach. *Plant Physiol.* 151, 4. doi: 10.1104/pp.109.136952
- Leveau, A., Reed, J., Qiao, X., Stephenson, M. J., Mugford, S. T., Melton, R. E., et al. (2019). Towards take-all control: a C-21 β oxidase required for acylation of triterpene defence compounds in oat. *New Phytol.* 221, 1. doi: 10.1111/nph.15456
- Liu, Q., Khakimov, B., Cardenas, P. D., Cozzi, F., Olsen, C. E., Jensen, K. R., et al. (2019). The cytochrome P450 CYP72A552 is key to production of hederagenin-based saponins that mediate plant defense against herbivores. *New Phytol.* 222, 3. doi: 10.1111/nph.15689
- Main, D., Cheng, C. H., Ficklin, S. P., Jung, S., Zheng, P., Coyne, C. J., et al. (2013). The cool season food legume database: an integrated resource for basic, translational and applied research, in *Plant and Animal Genome XXI Conference* (San Diego, CA).
- Marui, A., Nagafuchi, T., Shinogi, Y., Yasufuku, N., Omine, K., Kobayashi, T., et al. (2011). Cultivation research for high-glycyrrhizin licorice by applying low temperature and Ca²⁺ ion as environmental stress based on field investigation. *J. Fac. Agr. Kyushu Univ.* 56, 2.
- Matasci, N., Hung, L. H., Yan, Z., Carpenter, E. J., Wickett, N. J., Mirarab, S., et al. (2014). Data access for the 1,000 Plants (1KP) project. *Gigascience* 3, 1. doi: 10.1186/2047-217X-3-17
- Miettinen, K., Pollier, J., Buyst, D., Arendt, P., Csuk, R., Sommerwerk, S., et al. (2017). The ancient CYP716 family is a major contributor to the diversification of triterpenoid biosynthesis in eudicots. *Nat. Commun.* 8, 14153. doi: 10.1038/ncomms14153
- Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., et al. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* 30, 16. doi: 10.1002/jcc.21256
- Moses, T., Thevelein, J. M., Goossens, A., and Pollier, J. (2014). Comparative analysis of CYP93E proteins for improved microbial synthesis of plant triterpenoids. *Phytochemistry* 108, 47–56. doi: 10.1016/j.phytochem.2014.10.002
- Neamsuvan, O., Madeebing, N., Mah, L., and Lateh, W. (2015). A survey of medicinal plants for diabetes treating from Chana and Nathawee district, Songkhla province, Thailand. *J. Ethnopharmacol.* 174, 82–90. doi: 10.1016/j.jep.2015.07.050
- Osbourn, A. (1996). Saponins and plant defence—a soap story. *Trends Plant Sci.* 1, 1. doi: 10.1016/S1360-1385(96)80016-1
- Osbourn, A., Goss, R. J., and Field, R. A. (2011). The saponins-polar isoprenoids with important and diverse biological activities. *Nat. Prod. Rep.* 28, 1261–1268. doi: 10.1039/C1NP00015B
- Panchy, N., Lehti-Shiu, M., and Shiu, S. (2016). Evolution of gene duplication in plants. *Plant Physiol.* 171, 4. doi: 10.1104/pp.16.00523
- Prall, W., Hendy, O., and Thornton, L. E. (2016). Utility of a phylogenetic perspective in structural analysis of CYP72A enzymes from flowering plants. *PLoS One* 11, 9. doi: 10.1371/journal.pone.0163024
- Qi, X., Bakht, S., Qin, B., Leggett, M., Hemmings, A., Mellon, F., et al. (2006). A different function for a member of an ancient and highly conserved cytochrome P450 family: from essential sterols to plant defense. *Proc. Natl. Acad. Sci. U.S.A.* 103, 49. doi: 10.1073/pnas.0607849103
- Reed, J., Stephenson, M. J., Miettinen, K., Brouwer, B., Leveau, A., Brett, P., et al. (2017). A translational synthetic biology platform for rapid access to gram-scale quantities of novel drug-like molecules. *Metab. Eng.* 42, 185–192. doi: 10.1016/j.ymben.2017.06.012
- Saika, H., Horita, J., Taguchi-Shiobara, F., Nonaka, S., Nishizawa-Yokoi, A., Iwakami, S., et al. (2014). A novel rice cytochrome P450 gene, CYP72A31, confers tolerance to acetolactate synthase-inhibiting herbicides in rice and *Arabidopsis*. *Plant Physiol.* 166, 3. doi: 10.1104/pp.113.231266
- Sakai, H., Naito, K., Takahashi, Y., Sato, T., Yamamoto, T., Muto, I., et al. (2016). The *Vigna* Genome Server, 'VigGS': a genomic knowledge base of the genus *Vigna* based on high-quality, annotated genome sequence of the azuki bean, *Vigna angularis* (wild.) Ohwi & Ohashi. *Plant Cell Physiol.* 57, 1. doi: 10.1093/pcp/pcv189
- Schrödinger, L. L. C. (2017). *The PyMOL Molecular Graphics System, Version 2.0*.
- Seki, H., Ohyama, K., Sawai, S., Mizutani, M., Ohnishi, T., Sudo, H., et al. (2008). Licorice β -amyrin 11-oxidase, a cytochrome P450 with a key role in the biosynthesis of the triterpene sweetener glycyrrhizin. *Proc. Natl. Acad. Sci. U.S.A.* 105, 37. doi: 10.1073/pnas.0803876105
- Seki, H., Sawai, S., Ohyama, K., Mizutani, M., Ohnishi, T., Sudo, H., et al. (2011). Triterpene functional genomics in licorice for identification of CYP72A154 involved in the biosynthesis of glycyrrhizin. *Plant Cell* 23, 11. doi: 10.1105/tpc.110.082685
- Seki, H., Tamura, K., and Muranaka, T. (2015). P450s and UGTs: key players in the structural diversity of triterpenoid saponins. *Plant Cell Physiol.* 56, 8. doi: 10.1093/pcp/pcv062
- Shibata, S. (2000). A drug over the millennia: pharmacognosy, chemistry, and pharmacology of licorice. *Yakugaku Zasshi* 120, 10. doi: 10.1248/yakushi1947.120.10_849
- Suzuki, H., Fukushima, E. O., Umemoto, N., Ohyama, K., Seki, H., and Muranaka, T. (2018). Comparative analysis of CYP716A subfamily enzymes for the heterologous production of C-28 oxidized triterpenoids in transgenic yeast. *Plant Biotechnol.* 35, 2. doi: 10.5511/plantbiotechnology.18.0416a
- Tang, H., Krishnakumar, V., Bidwell, S., Rosen, B., Chan, A., Zhou, S., et al. (2014). An improved genome release (version Mt4.0) for the model

- legume *Medicago truncatula*. *BMC Genomics* 15, 312. doi: 10.1186/1471-2164-15-312
- Vo, N. N., Fukushima, E. O., and Muranaka, T. (2017). Structure and hemolytic activity relationships of triterpenoid saponins and sapogenins. *J. Nat. Med.* 71, 1. doi: 10.1007/s11418-016-1026-9
- Xue, Z., Tan, Z., Huang, A., Zhou, Y., Sun, J., Wang, X., et al. (2018). Identification of key amino acid residues determining product specificity of 2,3-oxidosqualene cyclase in *Oryza* species. *New Phytol.* 218, 3. doi: 10.1111/nph.15080
- Yano, R., Takagi, K., Takada, Y., Mukaiyama, K., Tsukamoto, C., Sayama, T., et al. (2017). Metabolic switching of astringent and beneficial triterpenoid saponins in soybean is achieved by a loss-of-function mutation in cytochrome P450 72A69. *Plant J.* 89, 3. doi: 10.1111/tpj.13403
- Yoshikawa, M., Morikawa, T., Nakano, K., Pongpiriyadacha, Y., Murakami, T., and Matsuda, H. (2002). Characterization of new sweet triterpene saponins from *Albizia myriophylla*. *J. Nat. Prod.* 65, 11. doi: 10.1021/np020220l
- Zhao, H. L., Cho, K., Ha, Y. W., Jeong, T., Lee, W. S., and Kim, Y. S. (2006). Cholesterol-lowering effect of platycodin D in hypercholesterolemic ICR mice. *Eur. J. Pharmacol.* 537, 1–3. doi: 10.1016/j.ejphar.2006.03.032
- Zhou, P., Silverstein, K. A. T., Ramaraj, T., Guhlin, J., Denny, R., Liu, J., et al. (2017). Exploring structural variation and gene family architecture with De Novo assemblies of 15 *Medicago* genomes. *BMC Genomics* 18, 261. doi: 10.1186/s12864-017-3654-1
- Zhu, M., Wang, C. X., Sun, W. T., Zhou, A. Q., Wang, Y., Zhang, G. L., et al. (2018). Boosting 11-oxo- β -amyrin and glycyrrhetic acid synthesis in *Saccharomyces cerevisiae* via pairing novel oxidation and reduction system from legume plants. *Metab. Eng.* 45, 43–45. doi: 10.1016/j.ymben.2017.11.009

Conflict of Interest: Authors HSu and SS were employed by company Tokiwa Phytochemical Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Fanani, Fukushima, Sawai, Tang, Ishimori, Sudo, Ohyama, Seki, Saito and Muranaka. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Evolution of Structural Diversity of Triterpenoids

Pablo D. Cárdenas[†], Aldo Almeida[†] and Søren Bak^{*}

Department of Plant and Environmental Science, University of Copenhagen, Frederiksberg, Denmark

OPEN ACCESS

Edited by:

Kazuki Saito,
RIKEN Center for Sustainable
Resource Science (CSRS),
Japan

Reviewed by:

Tetsuo Kushiro,
Meiji University, Japan
Tessa Moses,
University of Edinburgh,
United Kingdom

*Correspondence:

Søren Bak
bak@plen.ku.dk

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Plant Metabolism and
Chemodiversity,
a section of the journal
Frontiers in Plant Science

Received: 19 July 2019

Accepted: 01 November 2019

Published: 17 December 2019

Citation:

Cárdenas PD, Almeida A and
Bak S (2019) Evolution of Structural
Diversity of Triterpenoids.
Front. Plant Sci. 10:1523.
doi: 10.3389/fpls.2019.01523

Plants have evolved to produce a blend of specialized metabolites that serve functional roles in plant adaptation. Among them, triterpenoids are one of the largest subclasses of such specialized metabolites, with more than 14,000 known structures. They play a role in plant defense and development and have potential applications within food and pharma. Triterpenoids are cyclized from oxidized squalene precursors by oxidosqualene cyclases, creating more than 100 different cyclical triterpene scaffolds. This limited number of scaffolds is the first step towards creating the vast structural diversity of triterpenoids followed by extensive diversification, in particular, by oxygenation and glycosylation. Gene duplication, divergence, and selection are major forces that drive triterpenoid structural diversification. The triterpenoid biosynthetic genes can be organized in non-homologous gene clusters, such as in *Avena* spp., Cucurbitaceae and *Solanum* spp., or scattered along plant chromosomes as in *Barbarea vulgaris*. Paralogous genes organized as tandem repeats reflect the extended gene duplication activities in the evolutionary history of the triterpenoid saponin pathways, as seen in *B. vulgaris*. We review and discuss examples of convergent and divergent evolution in triterpenoid biosynthesis, and the apparent mechanisms occurring in plants that drive their increasing structural diversity within and across species. Using *B. vulgaris*' saponins as examples, we discuss the impact a single structural modification can have on the structure of a triterpenoid and how this affect its biological properties. These examples provide insight into how plants continuously evolve their specialized metabolome, opening the way to study uncharacterized triterpenoid biosynthetic pathways.

Keywords: triterpenoid saponins, structural diversity, convergent evolution, plant specialized metabolism, unlinked versus clustered pathways

INTRODUCTION

Plants have evolved the strategy of constantly diversifying the chemical structures they produce, evolving into an astonishing number of so-called plant specialized metabolites—of which the majority are thought to be involved in plant defense. The general conception has been that divergent evolutionary processes is the major driving force. However, biochemical and molecular knowledge of how these compounds have evolved has demonstrated that convergent evolution is surprisingly common (Pichersky and Lewinsohn, 2011).

Triterpenoids are one of the largest groups of plant specialized metabolites with over 14,000 structures known (Hamberger and Bak, 2013; <http://dnp.chemnetbase.com>). In the current model of triterpenoid biosynthesis, the isoprenoid precursor 2,3-oxidosqualene is cyclized by signature enzymes oxidosqualene cyclases (OSCs) to a number of cyclical triterpene scaffolds. Subsequently,

these structures are oxygenated by various cytochromes P450 (P450s) and finally glycosylated by UDP-glycosyltransferases (UGTs) (**Figure 1**). Plant genomes contain about a dozen OSC, about 250 P450s and about 110 UGTs, which are involved in multiple pathways. The sheer size of these gene families makes it a challenge to determine the specific genes coding for enzymes involved in generating the diversity of triterpenoids based on sequence alone. The OSC step constitutes the key branching point leading to biosynthesis of either sterols and steroidal saponins or to non-steroidal triterpenoids (e.g., triterpenoid saponins) (Thimmappa et al., 2014) (**Figure 1**). It is widely accepted that the cyclization of 2,3-oxidosqualene into sterols, occurs *via* the protosteryl cation and proceeds in a “chair-boat-chair” conformation. On the other hand, triterpenoid biosynthesis, *via* the dammarenyl cation, involves “chair-chair-chair” conformation of 2,3-oxidosqualene. In plants, more than 100 different triterpene scaffolds are derived from 2,3-oxidosqualene by OSCs, with lupeol and β -amyrin being among the most wide spread and studied regarding their biosynthesis and role in plants (Xu et al., 2004; Stephenson et al., 2019).

The increasing availability of genome sequencing of evolutionary distant plant species and in-depth studies of triterpenoid pathways in recent years, has revealed an emerging picture that triterpenoid pathways are evolving recurrently and thus represent an interesting case of both divergent and convergent evolution. Typically, it is the same gene families (OSCs, P450s, and UGTs) that are constantly recruited and afterwards expanded through gene duplications leading to tandem repeats. Here, we review the current knowledge about five triterpenoid scaffolds (and their derivatives) as examples of both divergent and convergent evolution in specialized metabolism and discuss the selection pressure on the genes, their organization, and how plants cope with the toxic triterpenoids they produce.

CONVERGENT EVOLUTION IN BIOSYNTHESIS OF TRITERPENOID SCAFFOLDS

Dammarenyl Cation-Derived Triterpenoids Have Arisen Multiple Times During Angiosperm Evolution

The majority of the known triterpenoids arise from the dammarenyl cation (Xiong et al., 2005). Resins from trees of the Dipterocarpaceae are known as dammar, hence the first triterpenoids isolated from dammar resins were coined dammaranediol I and II (Mills and Werner, 1955; Mills, 1956). Currently, the most commercially-valuable representatives of this class of compounds are the diverse ginsenosides. They can be obtained from mature six-year-old rhizomes of *Panax ginseng* and because of their anticancer activities some ginsenosides are used for chemotherapy treatment (Leung and Wong, 2010). Tansakul et al. (2006) were the first to clone a Dammarenyl-II synthase (PgDDS) and showed that this OSC from *P. ginseng* catalyzes the first committed step of ginsenoside biosynthesis. The only other DDS characterized thereafter

belongs to *Centella asiatica* (Kim et al., 2009). *P. ginseng* and *C. asiatica* belong to the Apiales and phylogenetic analysis showed that both DDS grouped in the same branch suggesting the DDS in these species evolved from a common ancestor. As DDSs have not been elucidated from the phylogenetically distant Dipterocarpaceae, future work on this family will shed some light on the evolutionary history of DDSs and could indicate if they arose from convergent or divergent evolution.

Lupeol and β -amyrin are prevalent pentacyclic triterpenoids derived from the dammarenyl cation and they are ubiquitously found in many different plant species (**Figure 1**). Nevertheless, phylogenetic analysis have shown that the genes producing these scaffolds group distinctively in different clades. Shibuya et al. (1999) first distinguished two clades of lupeol synthases in plants; one which is composed of specific lupeol synthases and another which is composed of multi-functional OSCs producing α -, β -amyrin, and lupeol (Thimmappa et al., 2014; Khakimov et al., 2015). Site-directed mutagenesis experiments have shown that a single amino acid replacement could convert a lupeol synthase into a β -amyrin synthase (and conversely), indicating the noticeable role of specific residues may have played in the evolution of OSC product specificity and generation of triterpenoid diversity (Kushiro et al., 1999; Kushiro et al., 2000). Furthermore, phylogenetic analysis of both monocot and dicot OSCs by Xue et al. (2012) and Augustin et al. (2011), additionally distinguish two distinctive clades of β -amyrin synthase in monocots and dicots.

Lupeol and β -amyrin can be present in plants as unmodified compounds typically found in resins or waxes (Szakiel et al., 2012) or they have a major role as precursors for other specialized triterpenoid metabolites, usually involved in plant defense and development. Lupeol is involved in nodule formation in *Lotus japonicus* through regulation of *ENOD40* gene expression (Delis et al., 2011). Lupeol is also part of the cuticular wax surface of castor bean plant (*Ricinus communis*) where it was suggested to have a physiological role in protection against dehydration (Guhling et al., 2006). Additionally, lupeol is the precursor for betulonic acid, a triterpenoid that accumulates in the bark of birch tree and has a potent anticancer activity (Pisha et al., 1995; Kumar et al., 2018). β -amyrin is a precursor of glycyrrhizin, a triterpenoid saponin found in the roots and stolons of liquorice (*Glycyrrhiza glabra*) well known for its pharmaceutical properties and as a natural sweetener (Hayashi et al., 2002). β -amyrin is also precursor for the antifungal saponin avenacin A-1 in oat and for the potent insect-feeding deterrent hederagenin glycosides in the Brassicaceae *Barbarea vulgaris* (Kuzina et al., 2009; Nielsen et al., 2010; Khakimov et al., 2016; Liu et al., 2019). Additionally, β -amyrin seems to play a role in root development in oat (Kemen et al., 2014) and in *Lotus japonicus* (Krokida et al., 2013), suggesting that triterpenoids like lupeol and β -amyrin are not exclusively involved in plant defense.

α -Onocerin—A Seco-Triterpenoid (*sensu lato*) That Evolved Convergently

Seco-triterpenoids are characterized by the absence of a C-C bond that would normally form one of the consecutive rings

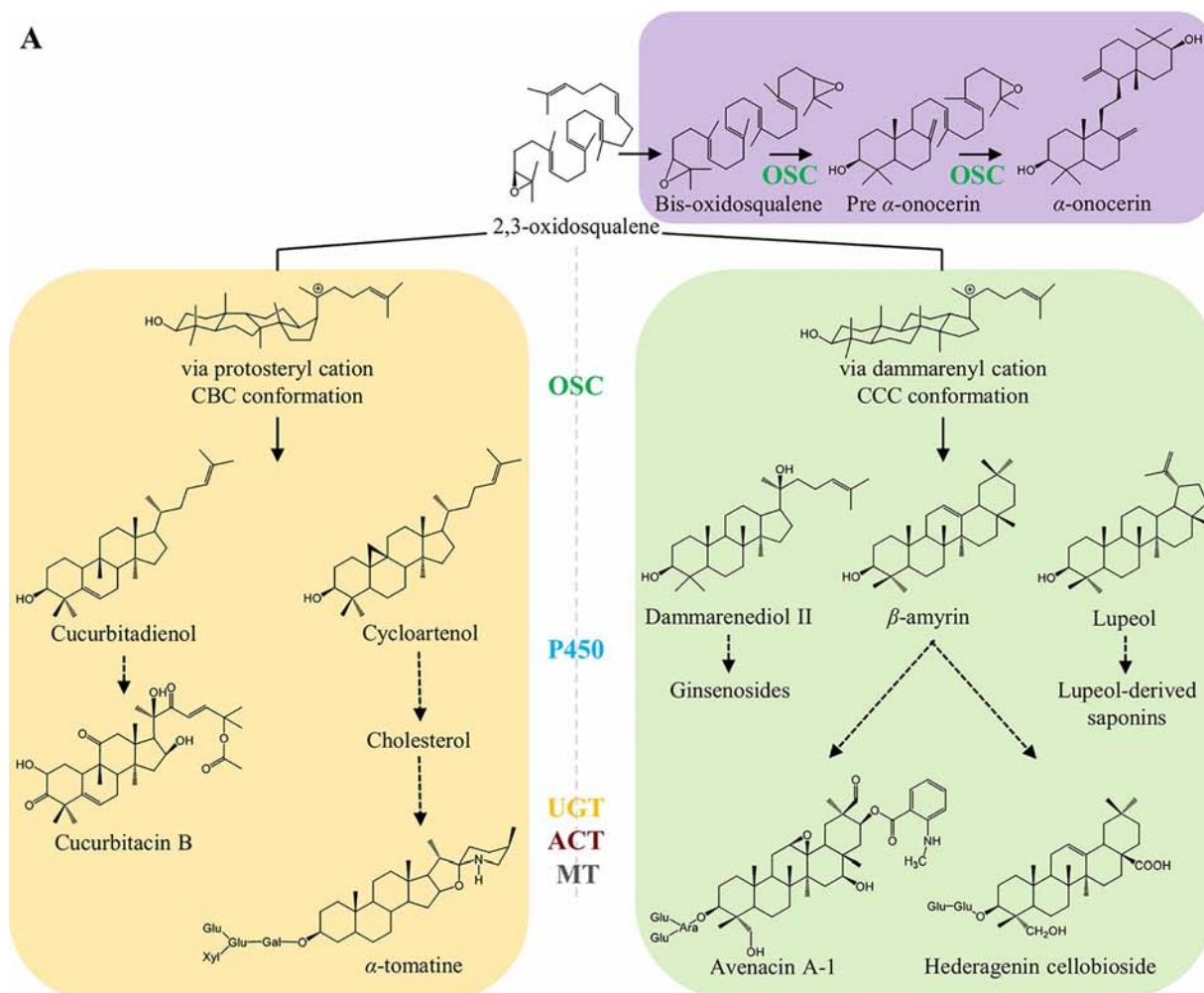
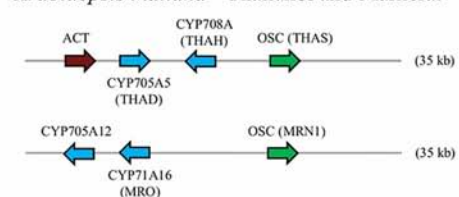
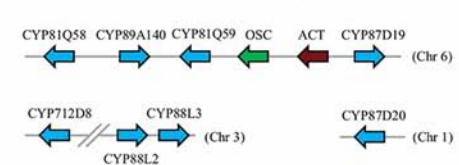
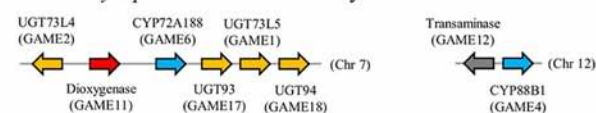
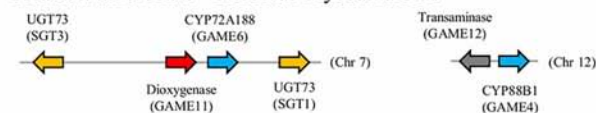
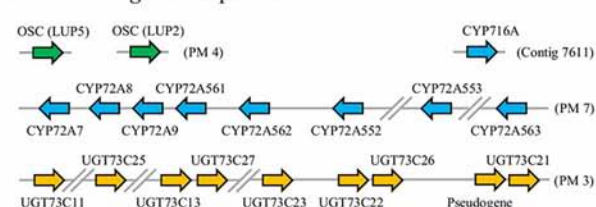
A**B***Avena* spp. - Avenacins*Arabidopsis thaliana* - Thalianol and Marneral*Cucumis sativus* - Cucurbitacins*Solanum lycopersicum* - Steroidal Glycoalkaloids*Solanum tuberosum* - Steroidal Glycoalkaloids*Barbarea vulgaris* - Saponins**FIGURE 1 |** Continued

FIGURE 1 | Simplified representation of the biosynthesis of sterols and triterpenoids in plants. **(A)** OSC signature enzymes catalyze the cyclization of 2,3-oxidosqualene, and in more rare cases bis-oxidosqualene, into several triterpenoid scaffolds. These structures can be further modified by tailoring enzymes, including oxygenation by P450s, glycosylation by UGTs, acylation by ACT, and methylation by MT. Selected structures are depicted and discussed in more detail in the text. Dashed arrows represent multiple biosynthetic reactions whereas solid arrows represent a single step. **(B)** Biosynthesis of plant triterpenoids can be mediated by non-homologous clustered genes or through non-linked genes. In *Avena* spp., a cluster of five genes are involved in the biosynthesis of avenacin A-1. In *Arabidopsis thaliana*, two clusters have been reported: thalianol cluster with four genes (up) and marnerial cluster with three genes (down). In *Cucumis sativus*, six genes associated with cucurbitacin biosynthesis are located in a cluster in chromosome 6, while four other genes are elsewhere in the genome. The core genes for biosynthesis of SGAs are clustered in chromosome 7 and 12 of *S. tuberosum* and *S. lycopersicum*, respectively. The key genes for biosynthesis of the insect-feeding deterrent hederagenin cellobioside are distributed along *B. vulgaris* genome in tandem repeats located at different pseudomolecules (PM). OSC, oxidosqualene cyclase; P450, cytochrome P450; UGT, UDP-glycosyltransferase; ACT, acyltransferase; MT, methyltransferase.

in a triterpenoid scaffold. They are known to be distributed across the plant kingdom. The *seco*-triterpenoid (*sensu lato*) α -onocerin consists of two bicyclic systems connected by a two-carbon linkage; its occurrence appears to be limited to Lycopod species and some species within the *Ononis* genus in the Fabaceae. Lycopods and the Fabaceae originated in very distant evolutionary times (Garratt et al., 1984; Giraud and Lejal-Nicol, 1989), which implies that the α -onocerin trait evolved convergently in Lycopods and in the *Ononis* genus. The biological function of α -onocerin still remains unknown.

On a biochemical level, α -onocerin biosynthesis differs from other triterpenoids as it is biosynthesized from 2,3;22,23-oxidosqualene (bis-oxidosqualene) instead of the typical triterpenoid precursor 2,3-oxidosqualene (Figure 1). In *Lycopodium clavatum*, biosynthesis of α -onocerin is carried out in two sequential steps by two paralogous OSCs (Araki et al., 2016). Pre- α -onocerin synthase (LcLCC), initiates cyclization from one of the epoxide bonds of bis-oxidosqualene. The cyclization terminates after formation of the A and B rings through the generation of pre- α -onocerin. Subsequently α -onocerin synthase (LcLCD) carries out the cyclization of the D and E rings from the remaining epoxide ring, to yield α -onocerin. Our group demonstrated that *Ononis spinosa* α -onocerin is biosynthesized by a single OSC (OsONS1) (Almeida et al., 2018). In *O. spinosa* a neofunctionalized squalene epoxidases (OsSQEs) provide the OSCs with the necessary bis-oxidosqualene. Fluorescence imaging microscopy experiments demonstrated protein-protein interactions between OsONS1 and the neofunctionalized OsSQEs (Almeida et al., 2018).

Phylogenetic analysis revealed that OsONS1 branches off from lupeol synthases, and is phylogenetically distant from the two *L. clavatum* α -onocerin synthases which branch off directly from sterol biosynthesis; thus, α -onocerin biosynthesis evolved convergently in these two plant species (Almeida et al., 2018). In addition, molecular docking simulation showed that OsONS1 alone produces α -onocerin in two cyclization steps, as opposed to the specialized LcLCC and LcLCD in Lycopods.

Protosteryl Cation-Derived Triterpenoids Are Found Across Angiosperm Phylogeny

Cucurbitacins are highly oxygenated tetracyclic triterpenes initially discovered in members of the Cucurbitaceae family and well known for their bitterness and toxicity (Metcalf et al., 1980). Cucurbitacins have been reported in 17 taxonomically distant related families in eudicots and some monocots. They consist of hundreds of derivatives from 20 main cucurbitacin molecules named from cucurbitacin A to T (Chen et al., 2005). At least 100 species in 30 genera of the Cucurbitaceae have been shown to contain cucurbitacins (Raemisch

and Turpin, 1984), with cucurbitacin B (Figure 1A) being present in 91% of the species (Chen et al., 2005).

Cucurbitacins are extremely toxic to mammals (David and Vallance, 1955) and function as feeding deterrents against several insects (Nielsen et al., 1977; Bigger and Chaney, 1998) but they have also been found to be feeding stimulant for leaf beetles belonging to the Luperini tribe of the Chrysomelidae (Metcalf, 1986). Cucurbitacin B can displace the insect steroidal hormone 20-hydroxyecdysone, thus affecting morphological changes in *Drosophila melanogaster* (Dinan et al., 1997).

The cucurbitacin biosynthesis pathway has mainly been studied in members of the Cucurbitaceae. The OSC cucurbitadienol synthase from *Cucurbita pepo* (CpCPQ) cyclizes 2,3-oxidosqualene to cucurbitadienol (Figure 1A) (Shibuya et al., 2004). The CpCPQ gene evolved after the divergence of dicots and monocots when the ancestral cycloartenol synthase gene duplicated, creating two clades of cycloartenol synthases (CASI and CASII) of which CpCPQ evolved from the *C. pepo* cycloartenol synthase (CpCPX) in the CASII clade (Xue et al., 2012).

Cucurbitacins are also present outside the Cucurbitaceae; for example *Iberis amara* belonging to the Brassicaceae contain cucurbitacins (Nielsen et al., 1977). The elucidation of the cucurbitacin biosynthetic pathway in *Iberis* will help to clarify the evolutionary history of Cucurbitacin biosynthesis.

Steroidal glycoalkaloids (SGAs) are triterpene-derived compounds found in major dicot Solanaceae crops such as tomato and potato. SGAs contain a nitrogen incorporated on their steroidal scaffold and they provide the plant with a barrier against a broad range of herbivores and pathogens, but are bitter and considered anti-nutritional compounds for humans (Cárdenas et al., 2015). SGAs and non-nitrogenous steroidal saponins are also present in distantly related monocot Liliaceae plant species. In these plant species, they have had special attention due to their pharmacological properties. For instance, the Liliaceae *Veratrum californicum* produces the potent anticancer molecule cyclopamine (Augustin et al., 2015). Biochemical and phylogenetic analysis of the enzymes involved in the biosynthesis of *Solanum* and *Veratrum* SGAs suggested their convergent origins and their partial recruitment from primary phytosterol metabolism (Augustin et al., 2015; Sonawane et al., 2016).

GENOME ORGANIZATION OF TRITERPENOID PATHWAYS

The biosynthetic genes for triterpenoids display different genome organizations across plant species (Figure 1B). While

in some species they are arranged in clusters of non-homologous genes (e.g., avenacins in oat, Qi et al., 2004; cucurbitacins in cucumber, Shang et al., 2014); the biosynthesis of other triterpenoids is mediated by genes scattered along the plant genome organized typically in tandem repeats (e.g., saponins in *Barbarea vulgaris*, Khakimov et al., 2015; Erthmann et al., 2018; Liu et al., 2019; and mogrosides in *Siraitia grosvenorii*, Itkin et al., 2016).

Clustered Genes Mediating Biosynthesis of Avenacins, Cucurbitacins and Steroidal Glycoalkaloids

Organization of genes in operon-like clusters has typically been associated with fungal genomes and operons present in bacteria. Nevertheless, nearly 20 metabolic gene clusters for plant specialized metabolites have been reported in multiple plant species (Boycheva et al., 2014). The avenacin, thalianol, and marnerial triterpenoid biosynthetic pathways were among the first to be reported being organized in non-homologous gene clusters in oat and *Arabidopsis thaliana*, respectively (Figure 1B, Qi et al., 2004; Field and Osbourn, 2008; Field et al., 2011). The organization of metabolic genes in clusters is thought to provide both co-inheritance, co-regulation and avoiding accumulation of toxic intermediates and may thus be beneficial for securing stable inheritance of functional chemical defense pathways in a dynamic ecological context of natural populations (Osborn, 2010; Takos and Rook, 2012).

Triterpenoid saponins are found mainly in dicotyledonous species. *Avena* spp. is the only known triterpenoid saponin producing monocotyledon. *Avena* spp. accumulates avenacins which are saponins produced in the roots of oat species and provide the plant with a potent barrier against soil-borne fungi (Papadopoulou et al., 2002) (Figure 1). SAD1, a β -amyrin synthase, catalyzes the first committed step in avenacin biosynthesis (Qi et al., 2004), and is present in the oat genome in an operon like cluster together with genes coding for the tailoring enzymes required for avenacin biosynthesis, including: SAD2, the P450 CYP51H10 that oxidizes β -amyrin; and three genes that act together in the acylation steps of avenacin: SAD9, a N-methyltransferase; SAD10, a UGT74H5 glycosyltransferase; and SAD7, a serine carboxypeptidase-like acyltransferase (Mugford et al., 2013). CYP51H10 (SAD2) evolved from a predisposition in sterol metabolism of the orthologous CYP51 catalyzing the conserved 14 α -demethylation reaction in sterol metabolism in eukaryotes (Geisler et al., 2013).

In Cucurbitaceae, part of the cucurbitacin pathway is similarly clustered; in cucumber the cucurbitadienol synthase is flanked by four P450s and an ACT gene out of which two P450s and the ACT have been functionally characterized (Shang et al., 2014). The P450s in this cluster belong to different subfamilies, indicating this cluster was not formed by tandem duplications but by genome reorganization. Later Zhou et al. (2016) uncovered the same conserved syntenic loci in melon and watermelon by comparative analyses of their genomes. While the core cluster is overall syntenically conserved, the CYP88A60 catalyzing the hydroxylation at position C19 and the CYP87D20 catalyzing

hydroxylations at C11 and C20 both lie outside of the non-homologous gene cluster.

In Solanaceae, the core genes required for the biosynthesis of steroidal glycoalkaloids are part of a metabolic operon-like gene cluster (Figure 1B). Itkin et al. (2013) showed that in tomato, six SGA genes are arranged in a cluster on chromosome 7 (UGT73L5: GAME1; UGT93: GAME17; UGT94: GAME18; UGT73L4: GAME2; Dioxygenase: GAME11; and CYP72A188: GAME6), whereas two other genes are located next to each other on chromosome 12 (CYP88B1: GAME4; and Transaminase: GAME12). Similarly in potato, four SGA-related genes are located in chromosome 7 (UGT73: SGT3; Dioxygenase: GAME11; CYP72A188: GAME6; and UGT73: SGT1) and two in chromosome 12 (Transaminase: GAME12; and CYP88B1: GAME4).

Tandem Repeats of Triterpenoid-Biosynthetic Genes and Triterpenoid Diversity in *B. vulgaris*, *Medicago truncatula* and *Glycyrrhiza uralensis*

Triterpenoid saponin biosynthesis has evolved recurrently in evolution and thus the organization of the genes may not be conserved. In recent years, the *Barbarea* genus has appeared as a unique plant model as it is the only genus in the agronomically important cabbage family (Brassicaceae) known to accumulate saponins (Khakimov et al., 2016). Some of these saponins (e.g., hederagenin cellobioside) are highly deterrent to *Brassica* specialist herbivores such as flea beetles (*Phyllotreta nemorum*) and the diamondback moth (*Plutella xylostella*) (Kuzina et al., 2009; Kuzina et al., 2011; Liu et al., 2019).

Genome (Byrne et al., 2017) and genetic (Kuzina et al., 2009; Kuzina et al., 2011) analysis showed that the genes mediating the biosynthesis of the deterrent triterpenoid saponins in *B. vulgaris* are not linked, but are present in tandem repeats spread along the genome (Figure 1B). The OSCs, P450s, and UGTs in the pathway are generally characterized to be rather substrate and product promiscuous enzymes which may facilitate that more than 49 different saponin structures can be generated in *B. vulgaris* with a limited number of genes (Khakimov et al., 2016). QTLs for flea beetle resistance and accumulation of saponins co-localize in *B. vulgaris* (Kuzina et al., 2009; Kuzina et al., 2011; Khakimov et al., 2015). Two unlinked QTLs containing two OSC (i.e., LUP2 and LUP5) and a tandem repeat of eight P450s (CYP72As), respectively, were identified. *In vitro* and *in planta* assays shown that LUP5 was preferentially expressed in the insect-deterrent *B. vulgaris* G-type. LUP5 leads to predominant accumulation of β -amyrin, the precursor for the deterrent hederagenin cellobioside. Conversely, LUP2 was preferentially expressed in the P-type plants (insect-susceptible), and produces mainly lupeol, the precursor for lupeol-derived saponins that appear not to be deterrent. Of the eight CYP72As, only CYP72A552 oxidizes oleanolic acid at the C23 position leading to the formation of insect deterrent hederagenin glycosides (Liu et al., 2019). CYP716A80 and CYP716A81 catalyze C28 carboxylation and two additional hydroxylation (Khakimov et al., 2015) while

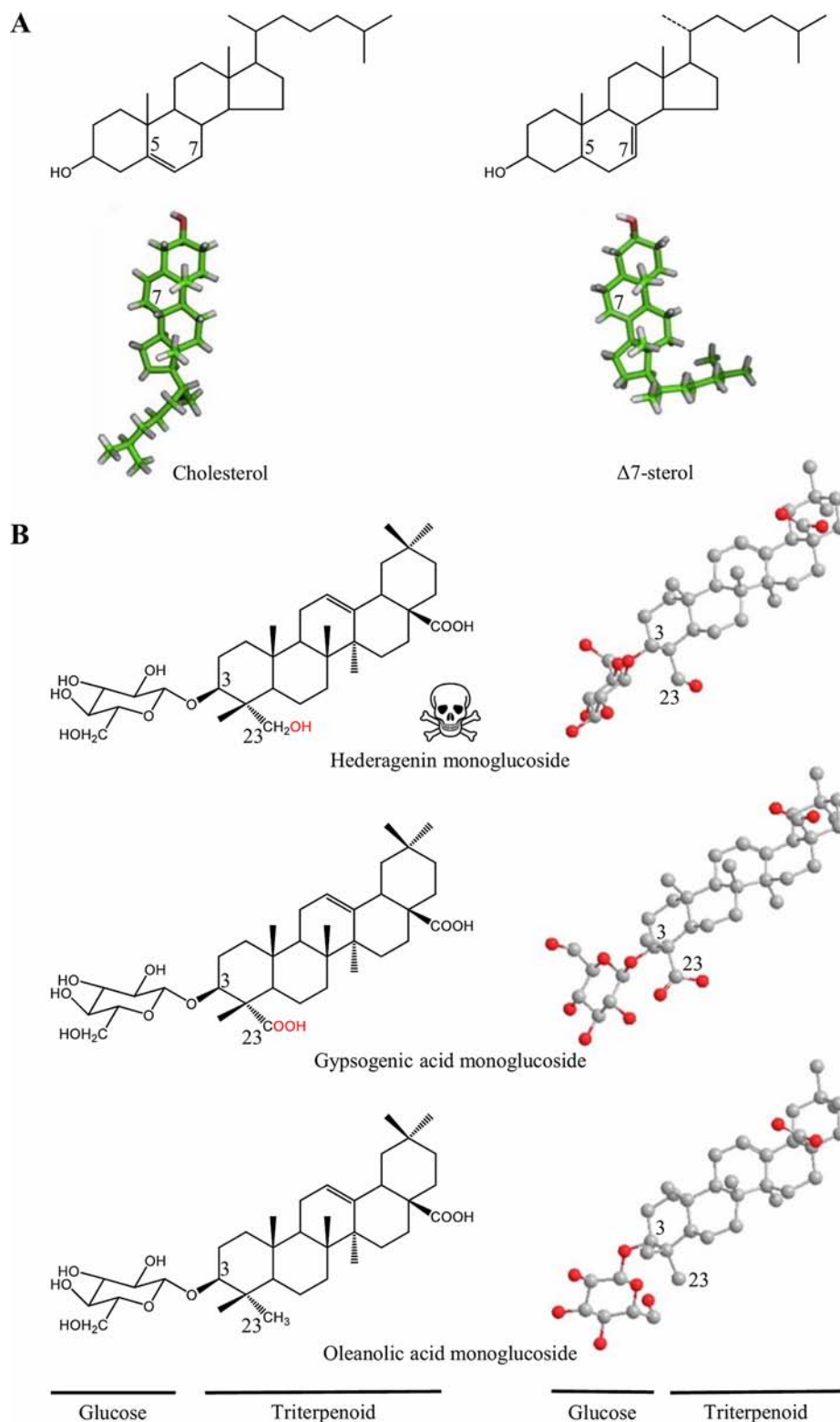


FIGURE 2 | Structural changes may affect the biological activity of triterpenoids. **(A)** Structural formulae (up) and 3D models (down) of cholesterol and Δ^7 sterol. In the Δ^7 sterol the side chain is bent as compared to cholesterol. In sea cucumber cell membrane systems, cholesterol are replaced by Δ^7 sterols to possible modulate the lytic action of saponins as the 3D structure is altered (Claereboudt et al., 2018). **(B)** Structural formulae (left) and 3D models (right) of *B. vulgaris* saponins. In the deterrent hederagenin monoglucoside, the glucose (C3) is twisted in respect to the triterpenoid backbone when a hydroxyl group is added at position C23 (Liu et al., 2019). 3D models are derived from Claereboudt et al. (2018) and Liu et al. (2019), respectively.

a tandem repeat of at least five UGT73Cs (Augustin et al., 2012; Erthmann et al., 2018) are also involved in the pathway. Both the CYP716As and the UGT73Cs are not linked to QTLs for neither insect resistance nor saponin accumulation. All members of the UGT73C tandem repeat accepted both hederagenin and oleanolic acid as substrates, but generated different glycosylated products, creating an spectrum of mono and bisdesmosidic saponins (Erthmann et al., 2018). If the *B. vulgaris* saponin non-linked gene organization is part of facilitating the opportunity to create multiple end products with a number of limited genes is an open question, but interestingly other pathways (e.g., anthocyanins and glucosinolates) also producing a multiple range of end products, are also known to be unlinked.

As in *B. vulgaris*, the genome of *Medicago truncatula* and the draft genome of *Glycyrrhiza uralensis* did not reveal operon-like gene cluster of OSCs, P450s, and UGTs involved in neither soyasaponins nor glycyrrhizin biosynthesis, but rather the gene candidates are organized as tandem repeats of each of these gene families (Naoumkina et al., 2010; Mochida et al., 2017). However, while most of the genes in these tandem repeats remain functionally uncharacterized, UGT73K1 was proven to glycosylate soyasapogenol B and E at the C3 position (Achnine et al., 2005).

WHAT DETERMINES TOXICITY VERSUS AUTOTOXICITY OF PLANT TRITERPENOID SAPONINS?

Saponins are generally assumed to target and disrupt vital functions common among organisms, such as the integrity of the cell membrane. In this regard, a fundamental paradox in plant evolution is that many defense metabolites may be harmful for the plant themselves. Therefore, plants have had to evolve strategies to both biosynthesize and accumulate defense molecules without causing autotoxicity. Some of the strategies used by plants to avoid autotoxicity are the compartmentalization of defense compounds in specialized structures and chemical modifications leading to decreased toxicity (i.e., glycosylation). For instance, avenacin A-1 is normally localized in the vacuoles of root epidermal cells. Mutants for the *sad3* glucosyltransferase accumulate partially glycosylated avenacin A-1, having root epidermis defects and affected saponin cellular distribution (Mylona et al., 2008), illustrating the detrimental effects of pathway intermediates.

Even though saponins are among the largest classes of natural products, their mode of action is not really understood—this may partly be attributed to the often complex mixtures of saponins in plants. Saponins have been shown to cause increased insect mortality, lower food intake, weight reduction, and developmental problems among others (Agerbirk et al., 2003; Christensen et al., 2019); and due to their amphiphilic nature, their most studied effect is regarding their membrane permeabilizing properties. Early studies using artificial lipid bilayers have shown that avenacin A-1 induces permeabilization in a cholesterol-dependent manner and requires the presence of an intact glycoside moiety in the C3 position of the

triterpenoid scaffold (Armah et al., 1999). The dependence of the permeabilizing activity on the presence of cholesterol in the membrane and the sugar moiety also was shown for the steroidal saponin digitonin and glycoalkaloids. For example, α -chaconine had strong lytic activity of cholesterol-containing lipid vesicles compared to α -solanine, which has the same aglycone structure but different types of glycosides (Nishikawa et al., 1984; Keukens et al., 1995; Keukens et al., 1996).

Consequently, a prominent role in the avoidance of autotoxicity can be attributed to modifications in the cell membrane composition and further modifications on the saponin structure. Saponins are also present in marine organisms such as sea stars and sea cucumbers (Yendo et al., 2014). Interestingly, recent studies on how sea cucumbers tolerate the toxic saponins they produce, suggest sea cucumbers have replaced cholesterol by other sterols (e.g., $\Delta 7$ and $\Delta 9$ sterols) in their membranes, presumably in order to modulate the lytic action of its own saponins (Popov, 2003). A detailed analysis of the sterols present in sea cucumbers showed that although cholesterol and $\Delta 7$ sterols have essentially the same chemical formula and molecular weight (Figure 2), the double bond present in $\Delta 7$ sterols has a dramatic effect on its 3D structure, possibly affecting its molecular interaction with sea cucumbers own saponins (Claereboudt et al., 2018). A similar effect on the 3D structure has been speculated to play a role in the biological activity of saponins from *B. vulgaris* (Figure 2). The hydroxylation of oleanolic acid at position C23 catalyzed by CYP72A552, leading to the formation of the highly-deterrent hederagenin causes a rotation of the monoglucoside at C3 of c. 90°, relative to the plane of the aglycone. When a carboxyl group is introduced instead as in gypsogenic acid, the glucose is situated about the same plane as in oleanolic acid. Gypsogenic and oleanolic acid glycosides are much less toxic to larvae of both *Manduca sexta* and *Plutella xylostella* than they are compared to hederagenin monoglucoside (Liu et al., 2019). Accordingly, the rotation of the sugar moiety in hederagenin glycosides may be responsible for changes on its physiochemical properties with membrane sterols, and this could lead to alter its biological role.

CONCLUDING REMARKS

Increasing characterization of biosynthetic pathways for plant specialized metabolites is revealing that convergent evolution is surprisingly common (Pichersky and Lewinsohn, 2011). Nevertheless, we lack a biological understanding of how identical classes of specialized metabolites evolved recurrently across lineages and even across Kingdoms of Life. Biochemical analysis of the enzymes combined with molecular phylogenetic analysis have until now given detailed insights of how the enzymes for these compounds might have evolved. Further studies are needed to elucidate which biological pressures drive the recurrent emergence of the same or similar specialized metabolites across lineages of plants.

Genome analysis have revealed that some triterpenoid pathways are unlinked while others are organized in operon like clusters. The driving forces behind the genome organization of triterpenoid pathways is currently not understood and difficult

to address experimentally. However, as more plant genomes become readily available this key question in evolution of triterpenoid diversity might become addressable.

AUTHOR CONTRIBUTIONS

All the authors conceived and wrote the manuscript.

REFERENCES

- Achnine, L., Huhman, D. V., Farag, M. A., Sumner, L. W., Blount, J. W., and Dixon, R. A. (2005). Genomics-based selection and functional characterization of triterpene glycosyltransferases from the model legume *Medicago truncatula*. *Plant J.* 41, 875–887. doi: 10.1111/j.1365-3113.2005.02344.x
- Agerbirk, N., Olsen, C. E., Bibby, B. M., Frandsen, H. O., Brown, L. D., Nielsen, J. K., et al. (2003). A saponin correlated with variable resistance of *Barbarea vulgaris* to the diamondback moth *Plutella xylostella*. *J. Chem. Ecol.* 29, 1417–1433. doi: 10.1023/a:1024217504445
- Almeida, A., Dong, L., Khakimov, B., Bassard, J. E., Moses, T., Lota, F., et al. (2018). A single oxidosqualene cyclase produces the seco-triterpenoid α -Onocerin. *Plant Physiol.* 176, 1469–1484. doi: 10.1104/pp.17.01369
- Araki, T., Saga, Y., Marugami, M., Otaka, J., Araya, H., Saito, K., et al. (2016). Onocerin biosynthesis requires two highly dedicated triterpene cyclases in a fern *Lycopodium clavatum*. *ChemBioChem* 17, 288–290. doi: 10.1002/cbic.201500663
- Armah, C. N., Mackie, A. R., Roy, C., Price, K., Osbourn, A. E., Bowyer, P., et al. (1999). The membrane-permeabilizing effect of avenacin A-1 involves the reorganization of bilayer cholesterol. *Biophys. J.* 76, 281–290. doi: 10.1016/S0006-3495(99)77196-1
- Augustin, J. M., Kuzina, V., Andersen, S. B., and Bak, S. (2011). Molecular activities, biosynthesis and evolution of triterpenoid saponins. *Phytochem.* 72, 435–457. doi: 10.1016/j.phytochem.2011.01.015
- Augustin, J. M., Drok, S., Shinoda, T., Sanmiya, K., Nielsen, J. K., Khakimov, B., et al. (2012). UDP-glycosyltransferases from the UGT73C subfamily in *Barbarea vulgaris* catalyze saponin 3-O-glucosylation in saponin-mediated insect resistance. *Plant Physiol.* 160, 1881–1895. doi: 10.1104/pp.112.202747
- Augustin, M. M., Ruzicka, D. R., Shukla, A. K., Augustin, J. M., Starks, C. M., O'Neil-Johnson, M., et al. (2015). Elucidating steroid alkaloid biosynthesis in *Veratrum californicum*: production of verazine in Sf9 cells. *Plant J.* 82, 991–1003. doi: 10.1111/tpj.12871
- Bigger, D. S., and Chaney, W. E. (1998). Effects of *Iberis umbellata* (Brassicaceae) on insect pests of cabbage and on potential biological control agents. *Environ. Entomol.* 27, 161–167. doi: 10.1093/ee/27.1.161
- Boycheva, S., Daviet, L., Wolfender, J. L., and Fitzpatrick, T. B. (2014). The rise of operon-like gene clusters in plants. *Trends Plant Sci.* 19, 447–459. doi: 10.1016/j.tplants.2014.01.013
- Byrne, S. L., Erthmann, P. Ø., Agerbirk, N., Bak, S., Hauser, T. P., Nagy, I., et al. (2017). The genome sequence of *Barbarea vulgaris* facilitates the study of ecological biochemistry. *Sci. Rep.* 7, 40728. doi: 10.1038/srep40728
- Cárdenas, P. D., Sonawane, P. D., Heinig, U., Bocobza, S. E., Burdman, S., and Aharoni, A. (2015). The bitter side of the nightshades: genomics drives discovery in Solanaceae steroidal alkaloid metabolism. *Phytochem.* 113, 24–32. doi: 10.1016/j.phytochem.2014.12.010
- Chen, J. C., Chiu, M. H., Nie, R. L., Cordell, G. A., and Qiu, S. X. (2005). Cucurbitacins and cucurbitane glycosides: structures and biological activities. *Nat. Prod. Rep.* 22, 386–399. doi: 10.1039/b418841c
- Christensen, S., Enge, S., Jensen, K. R., Müller, C., Kiær, L. P., Agerbirk, N., et al. (2019). Different herbivore responses to two co-occurring chemotypes of the wild crucifer *Barbarea vulgaris*. *Arthropod. Plant Interact.* 13, 19–30. doi: 10.1007/s11829-018-9633-x
- Claereboudt, E. J. S., Eeckhaut, I., Lins, L., and Deleu, M. (2018). How different sterols contribute to saponin tolerant plasma membranes in sea cucumbers. *Sci. Rep.* 8, 1–11. doi: 10.1038/s41598-018-29223-x
- David, A., and Vallance, D. K. (1955). Bitter principles of Cucurbitaceae. *J. Pharm. Pharmacol.* 7, 295–296. doi: 10.1097/00006247-197402000-00002
- Delis, C., Krokida, A., Georgiou, S., Pena-Rodriguez, L. M., Kavrroulakis, N., Ioannou, E., et al. (2011). Role of lipoel synthase in *Lotus japonicus* nodule formation. *New Phytol.* 335–346. doi: 10.1111/j.1469-8137.2010.03463.x
- Dinan, L., Whiting, P., Girault, J. P., Lafont, R., Dhadialla, T. S., Cress, D. E., et al. (1997). Cucurbitacins are insect steroid hormone antagonists acting at the ecdysteroid receptor. *Biochem. J.* 327, 643–650. doi: 10.1042/bj3270643
- Erthmann, P. Ø., Agerbirk, N., and Bak, S. (2018). A tandem array of UDP-glycosyltransferases from the UGT73C subfamily glycosylate saponins, forming a spectrum of mono- and bisdesmosidic saponins. *Plant Mol. Biol.* 97, 37–55. doi: 10.1007/s11103-018-0723-z
- Field, B., and Osbourn, A. E. (2008). Metabolic diversification-independent assembly of operon-like gene clusters in different plants. *Sci.* 320, 543–547. doi: 10.1126/science.1154990
- Field, B., Fiston-Lavier, A. S., Kemen, A., Geisler, K., Quesneville, H., and Osbourn, A. E. (2011). Formation of plant metabolic gene clusters within dynamic chromosomal regions. *Proc. Natl. Acad. Sci.* 108, 16116–16121. doi: 10.1111/j.1095-8339.1984.tb02566.x
- Garratt, M. J., Tims, J. D., Rickards, R. B., Chambers, T. C., and Douglas, J. G. (1984). The appearance of Baragwanathia (Lycophytina) in the Silurian. *Bot. J. Linn. Soc.* 89, 355–358. doi: 10.1111/j.1095-8339.1984.tb02566.x
- Geisler, K., Hughes, R. K., Sainsbury, F., Lomonosoff, G. P., Rejzek, M., Fairhurst, S., et al. (2013). Biochemical analysis of a multifunctional cytochrome P450 (CYP51) enzyme required for synthesis of antimicrobial triterpenes in plants. *Proc. Natl. Acad. Sci.* 110, E3360–E3367. doi: 10.1073/pnas.1309157110
- Giraud, B., and Lejal-Nicol, A. (1989). Cassinuin dongolense n. sp. bois fossile de caesalpinaceae du nubien du soudan septentrional. *Rev. Palaeobot. Palynol.* 59, 37–50. doi: 10.1016/0034-6667(89)90004-3
- Guhling, O., Hobl, B., Yeats, T., and Jetter, R. (2006). Cloning and characterization of a lipoel synthase involved in the synthesis of epicuticular wax crystals on stem and hypocotyl surfaces of *Ricinus communis*. *Arch. Biochem. Biophys.* 448, 60–72. doi: 10.1016/j.abb.2005.12.013
- Hamberger, B., and Bak, S. (2013). Plant P450s as versatile drivers for evolution of species-specific chemical diversity. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 368, 20120426. doi: 10.1098/rstb.2012.0426
- Hayashi, H., Huang, P., Kirakosyan, A., Inoue, K., Hiraoka, N., Ikeshiro, Y., et al. (2002). Cloning and characterization of a cDNA encoding β -amyrin synthase involved in glycyrrhizin and soyasaponin biosyntheses in licorice. *Biol. Pharm. Bull.* 24, 912–916. doi: 10.1248/bpb.24.912
- Itkin, M., Heinig, U., Tzfadia, O., Bhidi, A. J., Shinde, B., Cardenas, P. D., et al. (2013). Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. *Sci.* 341, 175–179. doi: 10.1126/science.1240230
- Itkin, M., Davidovich-Rikanati, R., Cohen, S., Portnoy, V., Doron-Faigenboim, A., Oren, E., et al. (2016). The biosynthetic pathway of the nonsugar, high-intensity sweetener mogroside V from *Siraitia grosvenorii*. *Proc. Natl. Acad. Sci.* 113, E7619–E7628. doi: 10.1073/pnas.1604828113
- Kemen, A. C., Honkanen, S., Melton, R. E., Findlay, K. C., Mugford, S. T., Hayashi, K., et al. (2014). Investigation of triterpene synthesis and regulation in oats reveals a role for β -amyrin in determining root epidermal cell patterning. *Proc. Natl. Acad. Sci.* 111, 8679–8684. doi: 10.1073/pnas.140153111
- Keukens, E. A. J., de Vrije, T., van den Boom, C., de Waard, P., Plasman, H. H., Thiel, F., et al. (1995). Molecular basis of glycoalkaloid induced membrane disruption. *Biochim. Biophys. Acta* 1240, 216–228. doi: 10.1016/0005-2736(95)00186-7

FUNDING

PC work was supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 752437. SB and AA were supported by grants from the Independent Research Fund Denmark grant No. 7017-00275B and Novo Nordisk Foundation grant No. NNF17OC0027646.

- Keukens, E. A., de Vrije, T., Jansen, L. A., de Boer, H., Janssen, M., de Kroon, A. I., et al. (1996). Glycoalkaloids selectively permeabilize cholesterol containing biomembranes. *Biochim. Biophys. Acta* 1279, 243–250. doi: 10.1016/0005-2736(95)00253-7
- Khakimov, B., Kuzina, V., Erthmann, P. Ø., Fukushima, E. O., Augustin, J. M., Olsen, C. E., et al. (2015). Identification and genome organization of saponin pathway genes from a wild crucifer, and their use for transient production of saponins in *Nicotiana benthamiana*. *Plant J.* 84, 478–490. doi: 10.1111/tpj.13012
- Khakimov, B., Tseng, L. H., Godejohann, M., Bak, S., and Engelsens, S. B. (2016). Screening for triterpenoid saponins in plants using hyphenated analytical platforms. *Mol.* 21, 1–19. doi: 10.3390/molecules21121614
- Kim, O. T., Lee, J. W., Bang, K. H., Kim, Y. C., Hyun, D. Y., Cha, S. W., et al. (2009). Characterization of a dammarenediol synthase in *Centella asiatica* (L.) Urban. *Plant Physiol. Biochem.* 47, 998–1002. doi: 10.1016/j.plaphy.2009.08.001
- Krokida, A., Delis, C., Geisler, K., Garagounis, C., Tsikou, D., Peña-Rodríguez, L. M., et al. (2013). A metabolic gene cluster in *Lotus japonicus* discloses novel enzyme functions and products in triterpene biosynthesis. *New Phytol.* 200, 675–690. doi: 10.1111/nph.12414
- Kumar, P., Bhaduria, A. S., Singh, A. K., and Saha, S. (2018). Betulinic acid as apoptosis activator: molecular mechanisms, mathematical modeling and chemical modifications. *Life Sci.* 209, 24–33. doi: 10.1016/j.lfs.2018.07.056
- Kushiro, T., Shibuya, M., and Ebizuka, Y. (1999). Chimeric triterpene synthase. A possible model for multifunctional triterpene synthase. *J. Am. Chem. Soc.* 121, 1208–1216. doi: 10.1021/ja983012h
- Kushiro, T., Shibuya, M., Masuda, K., and Ebizuka, Y. (2000). Mutational studies on triterpene synthases: engineering lupeol synthase into β -amyrin synthase. *J. Am. Chem. Soc.* 122, 6816–6824. doi: 10.1021/ja0010709
- Kuzina, V., Ekstrom, C. T., Andersen, S. B., Nielsen, J. K., Olsen, C. E., and Bak, S. (2009). Identification of defense compounds in *Barbarea vulgaris* against the herbivore *Phyllotreta nemorum* by an ecometabolomic approach. *Plant Physiol.* 151, 1977–1990. doi: 10.1104/pp.109.136952
- Kuzina, V., Nielsen, J. K., Augustin, J. M., Torp, A. M., Bak, S., and Andersen, S. B. (2011). *Barbarea vulgaris* linkage map and quantitative trait loci for saponins, glucosinolates, hairiness and resistance to the herbivore *Phyllotreta nemorum*. *Phytochem.* 72, 188–198. doi: 10.1016/j.phytochem.2010.11.007
- Leung, K. W., and Wong, A. S.-T. (2010). Pharmacology of ginsenosides: a literature review. *Chin. Med.* 5, 20. doi: 10.1186/1749-8546-5-20
- Liu, Q., Khakimov, B., Cardenas, P. D., Federico, C., Olsen, C. E., Jensen, K. R., et al. (2019). The cytochrome P450 CYP72A552 is key to production of hederagenin-based saponins that mediate plant defense against herbivores. *New Phytol.* 222, 1599–1609. doi: 10.1111/nph.15689
- Metcalfe, R. L., Metcalfe, R. A., and Rhodes, A. M. (1980). Cucurbitacins as kairomones for diabrotic beetle. *Proc. Natl. Acad. Sci. U. S. A.* 77, 3769–3772. doi: 10.1073/pnas.77.7.3769
- Metcalfe, R. L. (1986). Coevolutionary adaptations of rootworm beetles (Coleoptera: Chrysomelidae) to cucurbitacins. *J. Chem. Ecol.* 12, 1109–1124. doi: 10.1007/BF01638999
- Mills, J. S., and Werner, A. E. A. (1955). The chemistry of dammar resin. *J. Chem. Soc.* 3132–3140. doi: 10.1039/JR9550003132
- Mills, J. S. (1956). The constitution of the neutral, tetracyclic triterpenes of dammar resin. *J. Chem. Soc.* 2196–2202. doi: 10.1039/JR9560002196
- Mochida, K., Sakurai, T., Seki, H., Yoshida, T., Takahagi, K., Sawai, S., et al. (2017). Draft genome assembly and annotation of *Glycyrrhiza uralensis*, a medicinal legume. *Plant J.* 89, 181–194. doi: 10.1111/tpj.13385
- Mugford, S. T., Louveau, T., Melton, R., Qi, X., Bakht, S., Hill, L., et al. (2013). Modularity of plant metabolic gene clusters: a trio of linked genes that are collectively required for acylation of triterpenes in oat. *Plant Cell* 25, 1078–1092. doi: 10.1105/tpc.113.110551
- Mylona, P., Owatworakit, A., Papadopoulou, K., Jenner, H., Qin, B., Findlay, K., et al. (2008). Sad3 and Sad4 are required for saponin biosynthesis and root development in oat. *Plant Cell* 20, 201–212. doi: 10.1105/tpc.107.056531
- Naoumkina, M. A., Modolo, L. V., Huhman, D. V., Urbanczyk-Wochniak, E., Tang, Y., Sumner, L. W., et al. (2010). Genomic and coexpression analyses predict multiple genes involved in triterpene saponin biosynthesis in *Medicago truncatula*. *Plant Cell.* 22, 850–866. doi: 10.1105/tpc.109.073270
- Nielsen, J. K., Larsen, L. M., and Sørensen, H. (1977). Cucurbitacin E and I in *Iberis amara*: Feeding inhibitors for *Phyllotreta nemorum*. *Phytochem.* 16, 1519–1522. doi: 10.1016/0031-9422(77)84014-4
- Nielsen, J. K., Nagao, T., Okabe, H., and Shinoda, T. (2010). Resistance in the plant, *Barbarea vulgaris*, and counter-adaptations in flea beetles mediated by saponins. *J. Chem. Ecol.* 36, 277–285. doi: 10.1007/s10886-010-9758-6
- Nishikawa, M., Nojima, S., Akiyama, T., and Inoue, K. (1984). Interaction of digitonin and its analogs with membrane cholesterol. *J. Biochem.* 96, 1231–1239. doi: 10.1093/oxfordjournals.jbchem.a134941
- Osbourne, A. (2010). Gene clusters for secondary metabolic pathways: an emerging theme in plant biology. *Plant Physiol.* 154, 531–535. doi: 10.1104/pp.110.161315
- Papadopoulou, K., Melton, R. E., Leggett, M., Daniels, M. J., and Osbourne, A. E. (2002). Compromised disease resistance in saponin-deficient plants. *Proc. Natl. Acad. Sci.* 96, 12923–12928. doi: 10.1073/pnas.96.22.12923
- Pichersky, E., and Lewinsohn, E. (2011). Convergent evolution in plant specialized metabolism. *Annu. Rev. Plant Biol.* 62, 549–566. doi: 10.1146/annurev-arplant-042110-103814
- Pisha, E., Chai, H., Lee, I. S., Chagwedera, T. E., Farnsworth, N. R., Cordell, G. A., et al. (1995). Discovery of betulinic acid as a selective inhibitor of human melanoma that functions by induction of apoptosis. *Nat. Med.* 1, 369–375. doi: 10.1038/nm1095-1046
- Popov, A. M. (2003). Comparative study of effects of various sterols and triterpenoids on permeability of model lipid membranes. *J. Evol. Biochem. Physiol.* 39, 314–320. doi: 10.1023/A:1026147925642
- Qi, X., Bakht, S., Leggett, M., Maxwell, C., Melton, R., and Osbourne, A. (2004). A gene cluster for secondary metabolism in oat: implications for the evolution of metabolic diversity in plants. *Proc. Natl. Acad. Sci. U. S. A.* 101, 8233–8238. doi: 10.1073/pnas.0401301101
- Raemisch, D. R., and Turpin, F. T. (1984). Field tests for an adult western corn rootworm aggregation pheromone associated with the phagostimulatory characteristic of bitter *Cucurbita* spp. *J. Agric. Entomol.* 1, 339–344.
- Shang, Y., Ma, Y., Zhou, Y., Zhang, H., Duan, L., Chen, H., et al. (2014). Biosynthesis, regulation, and domestication of bitterness in cucumber. *Science* 346, 1084–1088. doi: 10.1126/science.1259215
- Shibuya, M., Zhang, H., Endo, A., Shishikura, K., Kushiro, T., and Ebizuka, Y. (1999). Two branches of the lupeol synthase gene in the molecular evolution of plant oxidosqualene cyclases. *Eur. J. Biochem.* 266, 302–307. doi: 10.1046/j.1432-1327.1999.00875.x
- Shibuya, M., Adachi, S., and Ebizuka, Y. (2004). Cucurbitadienol synthase, the first committed enzyme for cucurbitacin biosynthesis, is a distinct enzyme from cycloartenol synthase for phytosterol biosynthesis. *Tetrahedron* 60, 6995–7003. doi: 10.1016/j.tet.2004.04.088
- Sonawane, P. D., Pollier, J., Panda, S., Szymanski, J., Massalha, H., Yona, M., et al. (2016). Plant cholesterol biosynthetic pathway overlaps with phytosterol metabolism. *Nat. Plants* 3, 16205. doi: 10.1038/nplants.2016.205
- Stephenson, M. J., Field, R. A., and Osbourne, A. (2019). The protosteryl and dammarenyl cation dichotomy in polycyclic triterpene biosynthesis revisited: has this ‘rule’ finally been broken? *Nat. Prod. Rep.* 36, 1044–1052. doi: 10.1039/c8np00096d
- Szakiel, A., Pączkowski, C., Pensec, F., and Bertsch, C. (2012). Fruit cuticular waxes as a source of biologically active triterpenoids. *Phytochem. Rev.* 11, 263–284. doi: 10.1007/s11101-012-9241-9
- Takos, A. M., and Rook, F. (2012). Why biosynthetic genes for chemical defense compounds cluster. *Trends Plant Sci.* 17, 383–388. doi: 10.1016/j.tplants.2012.04.004
- Tansakul, P., Shibuya, M., Kushiro, T., and Ebizuka, Y. (2006). Dammarenediol-II synthase, the first dedicated enzyme for ginsenoside biosynthesis, in *Panax ginseng*. *FEBS Letters* 580, 5143–5149. doi: 10.1016/j.febslet.2006.08.044
- Thimmappa, R., Geisler, K., Louveau, T., O'Maille, P., and Osbourne, A. (2014). Triterpene biosynthesis in plants. *Annu. Rev. Plant Biol.* 65, 225–257. doi: 10.1146/annurev-arplant-050312-120229
- Xiong, Q., Rocco, F., Wilson, W. K., Xu, R., Ceruti, M., and Matsuda, S. P. R. (2005). Structure and reactivity of the dammarenyl cation: configurational transmission in triterpene synthesis. *J. Org. Chem.* 70, 5362–5375. doi: 10.1021/jo050147e

- Xu, R., Fazio, G. C., and Matsuda, S. P. T. (2004). On the origins of triterpenoid skeletal diversity. *Phytochem.* 65, 261–291. doi: 10.1016/j.phytochem.2003.11.014
- Xue, Z., Duan, L., Liu, D., Guo, J., Ge, S., Dicks, J., et al. (2012). Divergent evolution of oxidosqualene cyclases in plants. *New Phytol.* 193, 1022–1038. doi: 10.1111/j.1469-8137.2011.03997.x
- Yendo, A., Costa, F., Da Costa, C. T., Colling, L., Gosmann, G., and Fett-Neto, A. (2014). Biosynthesis of plant triterpenoid saponins: genes, enzymes and their regulation. *Mini Rev. Org. Chem.* 11, 292–306. doi: 10.2174/1570193x1103140915111425
- Zhou, Y., Ma, Y., Zeng, J., Duan, L., Xue, X., Wang, H., et al. (2016). Convergence and divergence of bitterness biosynthesis and regulation in Cucurbitaceae. *Nat. Plants* 2, 16183. doi: 10.1038/nplants.2016.183

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Cárdenas, Almeida and Bak. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Evolution of Flavonoid Biosynthesis: A Bryophyte Perspective

Kevin M. Davies^{1*}, Rubina Jibran¹, Yanfei Zhou¹, Nick W. Albert¹, David A. Brummell¹, Brian R. Jordan², John L. Bowman³ and Kathy E. Schwinn¹

¹ The New Zealand Institute for Plant and Food Research Limited, Palmerston North, New Zealand, ² Faculty of Agriculture and Life Sciences, Lincoln University, Christchurch, New Zealand, ³ School of Biological Sciences, Monash University, Melbourne, VIC, Australia

OPEN ACCESS

Edited by:

Kazuki Saito,
RIKEN Center for Sustainable
Resource Science (CSRS), Japan

Reviewed by:

Stefan Martens,
Fondazione Edmund Mach, Italy
Brenda S.J. Winkel,
Virginia Tech, United States

*Correspondence:

Kevin M. Davies
kevin.davies@plantandfood.co.nz

Specialty section:

This article was submitted to
Plant Metabolism
and Chemodiversity,
a section of the journal
Frontiers in Plant Science

Received: 29 October 2019

Accepted: 07 January 2020

Published: 04 February 2020

Citation:

Davies KM, Jibran R, Zhou Y,
Albert NW, Brummell DA,
Jordan BR, Bowman JL and
Schwinn KE (2020) The Evolution
of Flavonoid Biosynthesis: A
Bryophyte Perspective.
Front. Plant Sci. 11:7.
doi: 10.3389/fpls.2020.00007

The flavonoid pathway is one of the best characterized specialized metabolite pathways of plants. In angiosperms, the flavonoids have varied roles in assisting with tolerance to abiotic stress and are also key for signaling to pollinators and seed dispersal agents. The pathway is thought to be specific to land plants and to have arisen during the period of land colonization around 550–470 million years ago. In this review we consider current knowledge of the flavonoid pathway in the bryophytes, consisting of the liverworts, hornworts, and mosses. The pathway is less characterized for bryophytes than angiosperms, and the first genetic and molecular studies on bryophytes are finding both commonalities and significant differences in flavonoid biosynthesis and pathway regulation between angiosperms and bryophytes. This includes biosynthetic pathway branches specific to each plant group and the apparent complete absence of flavonoids from the hornworts.

Keywords: anthocyanin, auronidin, dirigent, polyphenol oxidase, transcription factor

OPENING COMMENTS

The flavonoid pathway is a core component enabling land plants to interact with their environment. Flavonoids have been demonstrated to assist in tolerance to both abiotic and biotic stresses, and their production can be induced by cold, UV-B light (UVB) or strong white light, nutrient deprivation, desiccation, salinity, metal toxicity, and pest and pathogen attack (Agati and Tattini, 2010; Cheynier et al., 2013; Landi et al., 2015; Davies et al., 2018). In angiosperms, flavonoids are also key for signaling to pollinators and seed dispersal agents. Flavonoids are formed within the larger phenylpropanoid pathway, starting with chalcones as the first flavonoids. During plant evolution, flavonoid pathway diversity has increased greatly, with more than 8,000 different structures reported from the relatively small number of plant species studied to date (Andersen and Markham, 2006). This multiplicity of structures and functions is thought to have assisted land plants to colonize the wide range of environments they now occupy. The over 8,000 compounds are grouped into relatively few classes of flavonoids, based on the core structure and/or biosynthetic origin. The major flavonoid classes are the flavones, flavonols, isoflavonoids, aurones, 3-deoxyanthocyanins, anthocyanins, proanthocyanidins (condensed tannins), and the recently reported “auronidins” (Berland et al., 2019) (**Figures 1 and 2**). However, there are also notable groups of related non-flavonoid compounds produced with the same starting precursors as used for chalcone formation, such as the stilbenes and bibenzyls. Most flavonoids are targeted to the vacuole

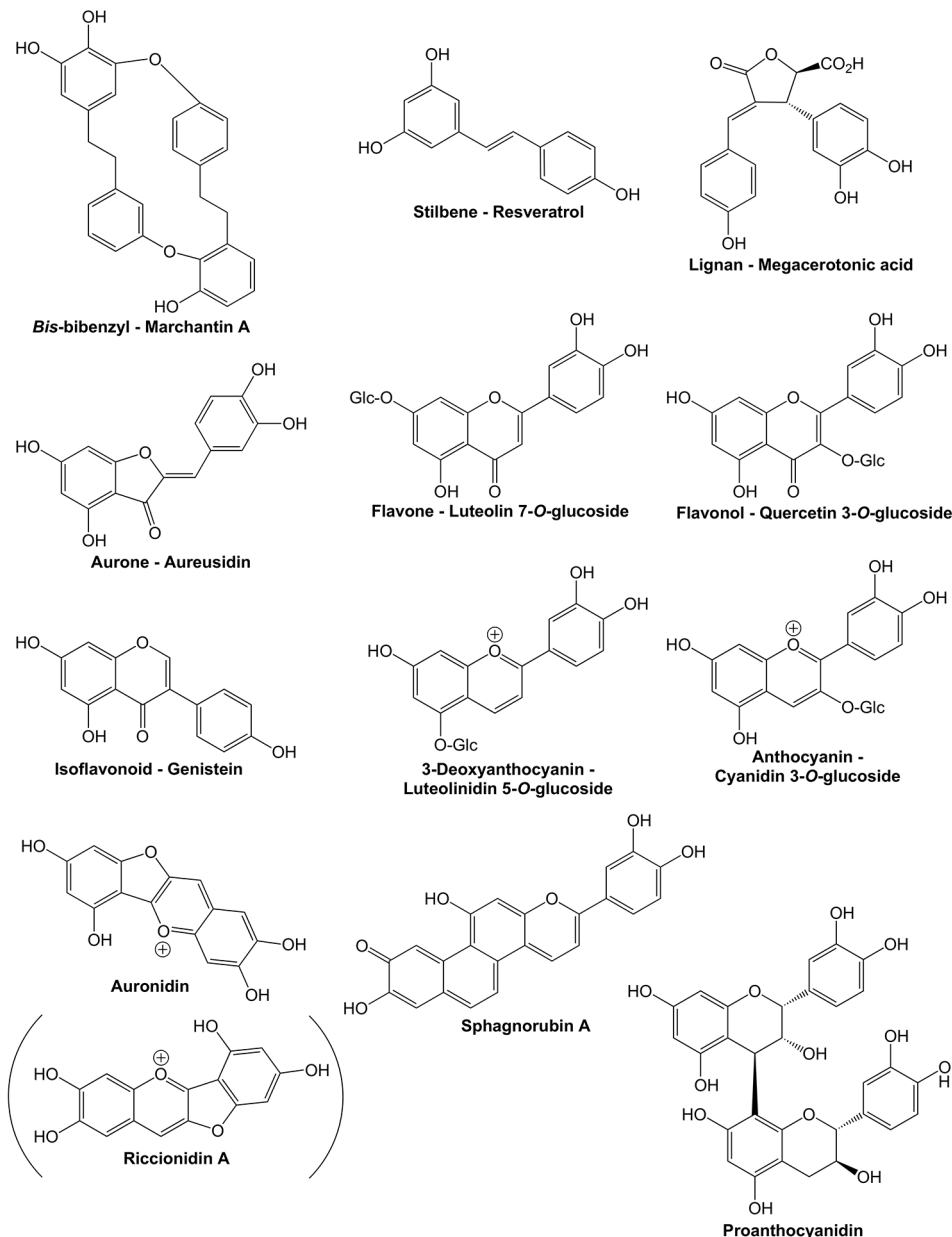


FIGURE 1 | Examples of compounds from phenylpropanoid pathway branches discussed in the text. Bibenzyls, stilbenes, and lignans are phenylpropanoids formed from branches prior to the start of the specific flavonoid pathway branch. The example compounds are from liverworts, angiosperms, and hornworts, respectively. All the other compounds shown are flavonoids.

as water-soluble glycosides, although some are transported to the cell wall or are released from the plant to the environment. Many can absorb light in the UV-spectrum, while anthocyanins and auronidins provide colored pigments that can screen in the visible part of the light spectrum (Lee and Gould, 2002; Landi et al., 2015; Berland et al., 2019).

Flavonoid biosynthesis is frequently considered unique to land plants. Some charophycean algae do tolerate terrestrial conditions (Karsten and Holzinger, 2014; Holzinger and Pichrtová, 2016) and can produce polyphenolics in response to

abiotic stresses such as UVB, salinity, or dehydration. Indeed, one hypothesis is that land plants arose from algal ancestors that were already terrestrial (Harholt et al., 2016) and therefore may have had some of the biosynthetic pathways characteristic of land plants. However, there are as yet no convincing reports of the flavonoid pathway branch existing outside of land plants, with the possible exception of some fungi. A common adaptation for UVB tolerance in extant algae is the production of mycosporine-like amino acids (MAA). However, while MAA production is frequently reported for red algae (Rhodophyta)

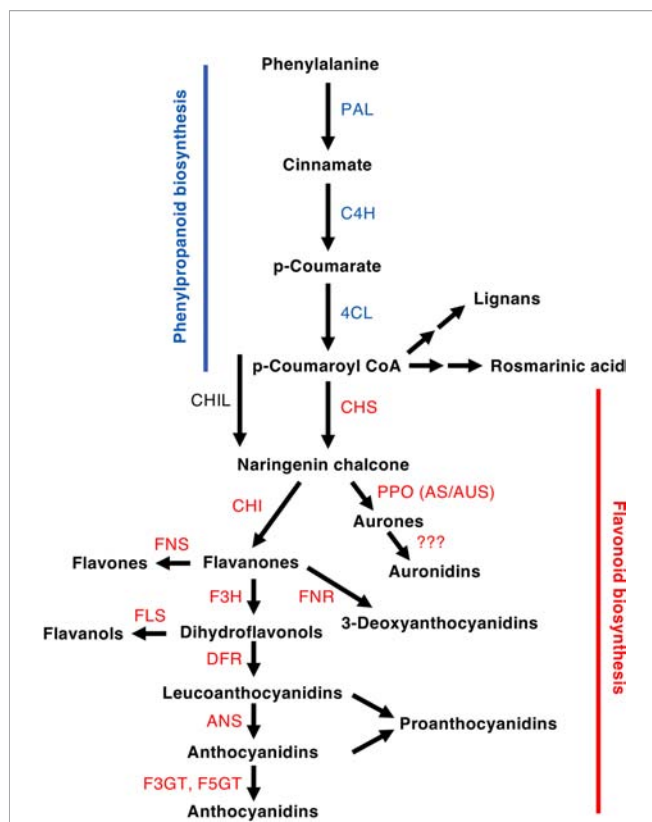


FIGURE 2 | Schematic of a section of the phenylpropanoid pathway leading to the production of major compound groups discussed in the article. Enzyme abbreviations are: PAL, phenylalanine ammonia-lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate:CoA ligase; CHS, chalcone synthase; CHI, chalcone isomerase; CHIL, chalcone isomerase-like; F3H, flavanone 3-hydroxylase; DFR, dihydroflavonol 4-reductase; FNR, flavanone reductase; ANS, anthocyanidin synthase; F3GT, flavonoid 3-O-glucosyltransferase; A5GT, anthocyanidin 5-O-glucosyltransferase; AUS/AS, aureusidin/aurone synthase; FNS, flavone synthase (2OGD = I, Cyp450 = II); FLS, flavonol synthase.

and other marine organisms, there are few reports for the Chlorophyta (a green algal clade that with Streptophyta form the Viridiplantae) and none for the charophyte algae that is thought to be most closely related to the land plant ancestor. Although genes for the initial steps of the phenylpropanoid pathway may be present in the genomes of extant Rhodophyta, Glaucophyta, Chlorophyta, and charophytes (Labeeuw et al., 2015; de Vries et al., 2017; Davies et al., 2020), there are no substantiated examples of flavonoid-specific genes being identified. Indeed, many metabolite studies on algae report only amounts of “total flavonoids” measured using general assays. Nevertheless, there are some reports detailing specific flavonoid structures in algal preparations, including compounds such as chalcones, flavones, flavonols, isoflavonoids, and proanthocyanidins (Klejduš et al., 2010; Goiris et al., 2014; El Shoubaky et al., 2016; Agregán et al., 2017; Ben Saad et al., 2017). Although the concentrations of flavonoid compounds reported in most examples are extremely low compared with those commonly found in land plants (ng gDW⁻¹ amounts

compared with mg gDW⁻¹), these reports mean that the presence of a biosynthetic pathway to flavonoids in algae cannot be ruled out.

The phenolics most commonly produced by algae are the phlorotannins, diverse oligomers derived from phloroglucinol that are found in brown algae (Heterokonts) (Imbs and Zvyagintseva, 2018). In *Ectocarpus siliculosus*, the phloroglucinol precursor is formed by the condensation of malonyl-CoA by a polyketide synthase (PKS) (Meslet-Cladière et al., 2013). This is analogous to the action of CHALCONE SYNTHASE (CHS) in the condensation of malonyl-CoA with a *p*-coumaroyl-CoA starter molecule for flavonoid biosynthesis. However, given the very large phylogenetic distance between brown algae and land plants this may be an example of parallel evolution. Notably, in fungal species there have been recent reports of both the production of flavonoids and the presence of genes with significant sequence similarity to those of the phenylpropanoid pathway of land plants. Findings include the detection of a range of phenylpropanoids, including flavonoids, in *Fusarium* (Bilska et al., 2018) and detection of flavonoids and candidate genes for stilbene production in *Alternaria* (Lu et al., 2019). The significance of phenylpropanoid biosynthesis being present in fungi for the current proposals on the evolutionary origins of flavonoid biosynthesis has yet to be addressed. The separation of the fungi and the algae/plant ancestors is thought to be an ancient event, preceding the divergence fungi and animals (Burki, 2014; Burki et al., 2020).

Origins and Vegetative Functions of the Flavonoid Pathway

Regardless of whether the ancestral genes for flavonoid biosynthesis were present in algal ancestors, the flavonoid pathway we see in extant land plants is hypothesized to have arisen when the land plant ancestors were first colonizing the land about 550–470 million years ago (MYA) (Markham, 1988; Stafford, 1991; Jorgensen, 1994; Koes et al., 1994; Kenrick and Crane, 1997; Rozema et al., 2002). Two major hypotheses have been proposed for the initial role of flavonoids. Firstly, that flavonoids may have helped in coping with the additional abiotic stresses resulting from a terrestrial lifestyle, in particular increased exposure to UVB, but potentially also drought and extreme temperature fluctuations (Markham, 1988; Jorgensen, 1994; Kenrick and Crane, 1997; Cockell and Knowland, 1999; Rozema et al., 2002; Ligrone et al., 2012; Mouradov and Spangenberg, 2014; Demarsy et al., 2017; Davies et al., 2018; de Vries and Archibald, 2018; Rensing, 2018). The alternative proposal is that flavonoids arose as physiological regulators or chemical messengers. This was outlined in Stafford (1991) for the regulation of auxin action, with signaling to mycorrhizal and symbiotic fungi proposed as possible additional communication functions. It was argued that flavonoids would probably have been present at only low concentrations when the pathway first evolved, limiting their efficacy as UVB-screening compounds. This was in the context of other arguments against the need for flavonoids as UVB-screens, such as the effective UVB-screening properties of non-flavonoid phenylpropanoids like the

hydroxycinnamic acids (HCAs). More recently, arguments for the early functions of flavonoids being other than UVB-screening have been extended by discoveries on their antioxidant properties and possible signaling actions through the redox pathway or by affecting H₂O₂ retrograde signals between the chloroplast and nucleus (Taylor and Grotewold, 2005; Agati and Tattini, 2010; Pollastri and Tattini, 2011; Agati et al., 2012; Brunetti et al., 2018; Foyer, 2018; Muhlemann et al., 2018; Brunetti et al., 2019).

That flavonoids affect auxin transport has now been demonstrated in several angiosperm species, including *Arabidopsis*, apple, and tomato, by analysis of mutants or transgenic lines with reduced flavonoid biosynthesis (Brown et al., 2001; Buer and Muday, 2004; Taylor and Grotewold, 2005; Peer and Murphy, 2007; Dare and Hellens, 2013; Maloney et al., 2014). Altered developmental traits in such plants include dwarfing, loss of pollen fertility and altered root development and gravitropic responses (Van der Meer et al., 1992; Napoli et al., 1999; Brown et al., 2001; Dare and Hellens, 2013; Maloney et al., 2014; Muhlemann et al., 2018). The phenotypes observed vary between species, for example the complete loss of flavonoids in the *Arabidopsis chs* mutant affects root patterning but not pollen viability (Burbulis et al., 1996; Ylstra et al., 1996). To date, the great majority of data are for angiosperms, and studies on other plant groups are required to determine whether these flavonoid functions are shared across land plants and so may have a common evolutionary origin in the early land plant ancestor. Differential distribution of flavonols and auxin has been observed accompanying stem reorientation in a gymnosperm (Ramos et al., 2016), supporting a conserved function in auxin transport within seed plants. If the genetic mechanisms involved are also conserved, then that would support an evolutionary origin before 350–300 MYA. However, while definitive experiments on flavonoids and hormone function have not been conducted in bryophytes, indications are that flavonoids are not necessary for normal development in this plant group. No flavonoids have been detected to date in hornworts, and a genetic mutant of the liverwort *Marchantia polymorpha* lacking flavonoids has normal developmental patterns (Clayton et al., 2018). Addition of phenylpropanoids or phenylpropanoid pathway inhibitors can alter bryophyte development in culture, but whether this is because of altered hormone action has not been tested (Chattopadhyay et al., 2018). Even with additional data it may be difficult to determine the most probable option between developmental roles for flavonoids having been acquired in seed plants since the last common ancestor, or having being present but then lost during subsequent evolution of the bryophytes. Nevertheless, establishing whether flavonoids regulate auxin action in bryophytes is an important goal.

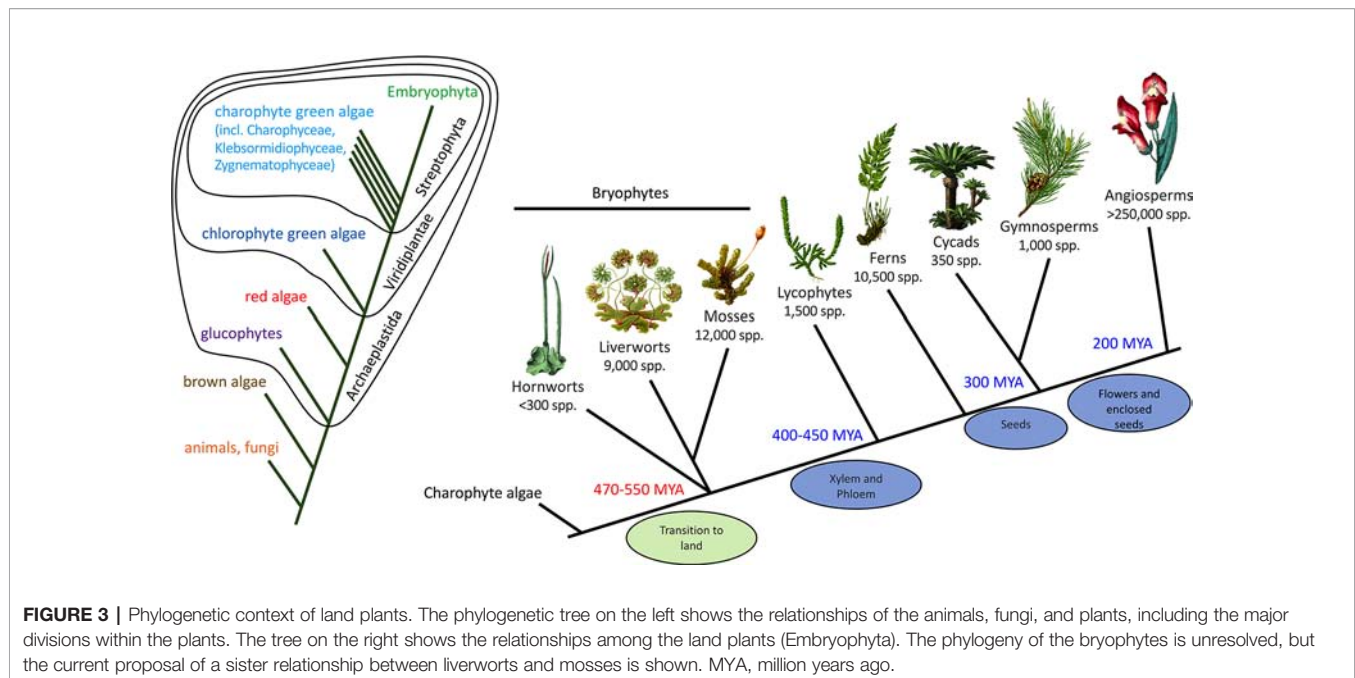
Fossils, such as those found in the Rhynie Chert in Scotland (about 410 MYA) (Ligrone et al., 2012), provide detail of the structure of early land plants but little information on their specialized/secondary metabolism. The presence of specialized biosynthetic or storage structures in fossils, such as possible equivalents to the terpenoid-accumulating oil bodies of extant

liverworts (Labandeira and Currano, 2013), can support the presence of specialized metabolite pathways but not provide details of the specific compounds produced. To generate hypotheses on the origins and subsequent evolution of the flavonoid pathway we need to compare the genetics and biochemistry of the pathway across diverse extant plant groups, as this can identify conserved pathway components that may have originated with the last common ancestor. In this respect, bryophytes are of key importance (**Figure 3**). “Bryophytes” is the collective name for non-vascular land plants, comprising the liverworts (Marchantiophyta, approximately 9,000 species), hornworts (Anthocerotophyta, approximately 300 species) and mosses (Bryophyta, approximately 12,000 species). Evidence such as morphological comparison with the fossil record has placed liverworts as the “sister” group to extant land plants—that is, at the base of the land plant evolutionary tree, making bryophytes paraphyletic. However, DNA sequencing data have suggested alternatives, in particular, either paraphyletic bryophytes with hornworts as a sister group to all other land plants, or a single monophyletic bryophyte clade that is sister to the vascular plants (Wickett et al., 2014; Puttick et al., 2018). It is generally accepted that land plants evolved from an ancestral charophyte, and the extant algal sister group is probably the order Zygnematales or a clade of the Zygnematales and Coleochaetales together (Zhong et al., 2014; Delwiche and Cooper, 2015; de Vries and Archibald, 2018).

Flavonoids and Tolerance to Ultraviolet B Light

The origin of the flavonoid pathway for providing tolerance to UVB has been supported by recent studies on the liverwort species *Marchantia*. (“*Marchantia*” is used in this article to refer to *M. polymorpha* subsp. *ruderalis*, which is the model experimental species.) *Marchantia* is an excellent research model: it is small; has a rapid growth rate; can asexually reproduce in large numbers through single-cell-derived clonal gemmae; and, has a small genome (approximately 220 Mb) which, although larger than that of *Arabidopsis* (at 135 Mb), contains significantly fewer genes (around 19,000 gene models compared with around 28,000 protein-coding genes in *Arabidopsis*) (Ishizaki et al., 2015; Bowman et al., 2016; Shimamura, 2016; Bowman et al., 2017). It also offers efficient CRISPR/Cas9 mutagenesis in a dominant haploid gametophytic generation (Sugano et al., 2018).

Significantly, the UVB response of *Marchantia* has many components in common with that of *Arabidopsis*. Flavonol O-glycosides are key for UVB tolerance of *Arabidopsis* (Kusano et al., 2011; Morales et al., 2013; Yin and Ulm, 2017), while the related flavone O-glycosides contribute to *Marchantia* UVB tolerance (Clayton et al., 2018). In both *Arabidopsis* and *Marchantia*, mutants with reduced flavonoid production are more easily damaged by UVB, while mutants or transgenics with increased flavonoid content have increased UVB tolerance. Moreover, the signaling pathway for flavonoid pathway induction through the UV RESISTANCE LOCUS8 (UVR8) photoreceptor, the bZIP transcription factor (TF) ELONGATED HYPOCOTYL5 (HY5),



and the modifiers of protein stability such as CONSTITUTIVELY PHOTOMORPHOGENIC1 (COP1) and REPRESSOR OF UVB PHOTOMORPHOGENESIS1 (RUP1) is also conserved between the species (Clayton et al., 2018; Kondou et al., 2019). Equivalent functional studies are lacking on other major basal plant groups, such as mosses and lycophytes. However, phenolics do seem to be important for UVB tolerance in mosses (Clarke and Robinson, 2008; Wolf et al., 2010; Waterman et al., 2017; Soriano et al., 2018b); and UVB exposure of the Antarctic moss *Pohlia nutans* increased transcript abundance for genes of the UVR8 and flavonoid pathways (Li et al., 2019). This suggests that the core UVB protection mechanism of UVR8-induced flavonoid production may already have been established in the last common ancestor of bryophytes and angiosperms. While some of the same genetic components, such as *UVR8* and *HY5*-like genes have been identified in distantly related groups within the Viridiplantae (Allorent et al., 2016; Bowman et al., 2017; Clayton et al., 2018; Soriano et al., 2018a; Kondou et al., 2019), and algae can produce purple phenolic pigments in response to abiotic stress (Aigner et al., 2013; Holzinger and Pichrtová, 2016), UVB-induced flavonoid production has not been characterized in algae. Thus, the UVR8-induction pathway for UVB-absorbing flavonoids (potentially flavone glycosides as the first compounds) may have been a character rapidly acquired during the water-to-land transition.

There are variations among land plants to the UVR8/flavonoid system for providing tolerance to UVB exposure, and further research would be beneficial to establish whether these may also have an early evolutionary origin. In *Arabidopsis*, the HCA compounds sinapate esters have an important role in UVB-screening. Comparison of *Arabidopsis* mutants at different biosynthetic steps found sinapate esters have a protective role comparable to, or perhaps more important

than, that of flavonols (Li et al., 1993; Landry et al., 1995). Furthermore, a genetic screen for UVB-tolerance genes identified the transcriptional repressor *AtMYB4* as being downregulated in response to UVB to facilitate increased sinapate ester production (Jin et al., 2000). The principal absorption maxima from 290 to 330 nm makes HCAs particularly effective UVB screening compounds, and would be a more carbon-efficient screen than the flavonoids. In addition to flavonol glycosides, mosses have been shown to produce biflavonoids or cell wall-bound phenolics in response to UVB exposure (Clarke and Robinson, 2008; Wolf et al., 2010; Waterman et al., 2017; Soriano et al., 2018b). Biflavonoid induction is a notable feature of the moss *Ceratodon purpureus*, a species found from Antarctica to hot desert environments (Waterman et al., 2017). The production of cell wall-localized phenolics as part of the UVB screening capacity may be a common feature of bryophytes, including for the flavonoid-lacking hornworts (Monforte et al., 2018; Soriano et al., 2018c).

Pigmented Flavonoids and Tolerance to Abiotic Stress

The other major group of flavonoids shown to be involved in tolerance to abiotic stress is the 3-hydroxyanthocyanins (typically cyanidin-derivatives) that are found in gymnosperms and almost all angiosperms (Lee and Gould, 2002; Landi et al., 2015). Under the physiological conditions commonly found in vegetative tissues, these provide red pigmentation. The structurally similar 3-deoxyanthocyanins have been extensively characterized in ferns, and also reported for lycophytes and mosses (Andersen and Markham, 2006). Although the color of 3-deoxyanthocyanins is shifted toward orange compared to the equivalent 3-hydroxyanthocyanins, in vegetative tissues similar red pigmentation typically occurs from both compound types.

Additionally, cell wall-bound red flavonoids have been reported from liverworts and mosses: riccionidin (an auronidin) and sphagnorubin, respectively (Vowinkel, 1975; Kunz et al., 1993; Berland et al., 2019). Until recently, riccionidin and sphagnorubin were considered anthocyanidins (the non-glycosylated anthocyanin core molecule) with additional rings. Thus, an evolutionary path could be envisioned of cell-wall bound anthocyanidin being the basal state that could have been present in the last common ancestor, then a progression to 3-deoxyanthocyanins, and then 3-hydroxyanthocyanins. However, now we know that riccionidins represent a separate flavonoid class unrelated to anthocyanins (see later sections for biosynthetic details) it is less clear as to what the class of red flavonoid pigment (if any) was present in the early land plants. Auronidin and anthocyanin biosynthesis may be derived characters in each lineage, or the missing pathway in each lineage may have been lost during evolution. Furthermore, no proanthocyanidins have been reported from bryophytes or lycophytes, suggesting this branch of flavonoid biosynthesis probably arose later in vascular plant evolution, since proanthocyanidins are found in extant ferns, gymnosperms, and angiosperms.

With regard to function, given the lack of flowers or enclosed seeds in the early land plants, any initial functions of pigmented flavonoids were probably unrelated to animal interactions. In vascular plants anthocyanins have been linked to improving tolerance to a range of abiotic stresses, as well as some biotic challenges. However, how anthocyanins achieve this in the different stress situations, and whether there is a single mechanism or specific functional variations, is the subject of much research and debate. Anthocyanins, auronidins and sphagnorubins can screen out white light to reduce photooxidative damage, and it has been estimated that anthocyanins could absorb over 40% of photosynthetically active radiation in the range containing the most damaging wavelengths for photoinhibition (Merzlyak and Chivkunova, 2000; Pietrini et al., 2002). However, anthocyanins can simultaneously reduce the cellular stress associated with photooxidation through quenching reactive oxygen species (ROS). The relative importance of these two mechanisms is unresolved, even for extensively studied situations such as the appearance of red anthocyanins during autumn senescence of leaves of deciduous trees. Localization of the pigmented flavonoids in the cell wall and vacuole could be thought to argue for a screening mechanism, as ROS are principally generated in plastids and mitochondria. However, it has been suggested that this does not rule out an antioxidant primary function (Agati et al., 2012). Many of the same arguments for or against light screening *versus* antioxidant primary functions can also be applied to the function of flavones/flavonols in UVB tolerance (Agati and Tattini, 2010; Agati et al., 2012; Davies et al., 2018). Further complicating the development of a unifying theory for anthocyanin abiotic stress function are alternative hypotheses that involve neither light screening nor ROS scavenging, such as drought tolerance through decreased osmotic potential, increasing light absorption to help warm

leaves, providing camouflage against insect herbivores, “honest” signaling to herbivores that leaves contain anti-feedant compounds and/or are about to be shed, and making leaves more noticeable to insect predators (anti-crypsis) (Gould et al., 1995; Lee and Gould, 2002; Manetas, 2006; Archetti, 2009; Hughes, 2011; Agati et al., 2012; Landi et al., 2015; Davies et al., 2018). Additionally, as mentioned earlier, there is also evidence supporting flavonoid roles as signaling molecules (Taylor and Grotewold, 2005; Agati and Tattini, 2010; Agati et al., 2012; Foyer, 2018).

The cell wall-bound nature of auronidins and sphagnorubins complicates the theory of an antioxidant or signaling role for red flavonoids. The cell wall-localization of the red pigments of mosses and liverworts, termed “tissue fixed”, has been known from early studies on *Sphagnum* and various liverworts (Nagai, 1915; Rudolph et al., 1977). Nevertheless, that the environmental stimuli inducing cell wall-bound red pigments in liverworts and mosses appear similar to those that trigger anthocyanin production in vascular plants, was also noted as far back as 1915 (Nagai, 1915). Kny (1890) and Cesares Gil (1902) noted that *M. polymorpha* or *Reboulia hemisphaerica*, respectively, grown in sunny locations produced more pigment than those growing in shady environments. Nagai (1915) was able to show that limiting nitrogen and phosphorus supply intensified the pigmentation of *M. polymorpha* and *M. paleacea*, but upon transfer to nutrient-rich media the newly developed tissue lacked pigmentation. Moreover, combinations of stresses that can cause oxidative stress are strong signals for red flavonoid biosynthesis in many bryophytes. Thus, as with angiosperms, in liverworts and mosses cold and light can individually induce reddening, but strong sunlight in cold conditions induces much stronger pigmentation, whether this be at altitude, in the Antarctic, or during the cold nights and bright days of autumn (Gerdol, 1996; Gerdol et al., 1998; Newsham et al., 2005; Hooijmaijers and Gould, 2007; Glime, 2007; Bonnett et al., 2010).

Detailed studies on *Marchantia* (Albert et al., 2018; Kubo et al., 2018) and *Ricciocarpos natans* (Kunz and Becker, 1995) found nitrogen deprivation and increased white light exposure induced auronidin accumulation, as has also been shown for anthocyanin accumulation in *Arabidopsis* and apple (Rubin et al., 2009; Wang et al., 2018). The signaling pathways for anthocyanin induction by nitrogen and phosphorus deficiency are well-characterized for *Arabidopsis*, with R2R3MYBs being the key activating transcription factors (Lillo et al., 2008; Rubin et al., 2009). Induction of auronidin in *Marchantia* by nitrogen and phosphorus also requires an R2R3MYB (Kubo et al., 2018), suggesting signaling components may be conserved. For Antarctic liverworts and mosses UVB exposure also induced production of red flavonoids, which most commonly were cell wall-bound (Newsham et al., 2005; Waterman et al., 2018). UVB induces anthocyanin production in some angiosperms, but it is much less common a response than induction of flavones/flavonols. Flavones and flavonols are more effective at screening UVB than anthocyanins, although aromatic acylation can give anthocyanins absorbance maxima in the UV range. The induction of anthocyanins by UVB has thus been suggested to be

more for ROS scavenging and/or screening of white light than for UVB screening. In the case of the non-acylated cell wall-bound flavonoid pigments of mosses and liverworts, it seems probable that production is induced to screen white light and prevent further ROS generation, especially as the summer conditions in the Antarctic present a combination of stresses from continuous white light, cold, and drought.

There are other red/purple plant pigments besides the flavonoids able to screen in photosynthetically active wavelengths. Notable among these are the betacyanins, which are produced in the many species of the core Caryophyllales that do not produce anthocyanins (Polturak and Aharoni, 2019), and the phenolic pigments of algae. *Zygonium ericetorum* is a charophyte green alga that can grow in alpine environments and when exposed to abiotic stress produces vacuolar-localized purple pigments, thought to be polymers of glucose and gallic acid, which can absorb in both UVB and photosynthetically active wavelengths (Aigner et al., 2013). In brown algae, phlorotannins can accumulate to more than 15% of dry weight (Imbs and Zvyagintseva, 2018). Phlorotannins are highly hydrophilic polymers, and may be cell wall-bound, stored intracellularly in vesicles, or exported.

Progress on determining the biological roles of cell wall-bound pigments in bryophytes has been limited by the lack of genetic systems, and the difficulty of extracting the pigments. However, genetic tools are now available in *Marchantia* that will allow tests of the functions of the pigments in abiotic or biotic stress tolerance. Mutants are available that have loss of auronidin pigmentation but retain flavone production (Albert et al., 2018; Kubo et al., 2018), have loss of flavone production but retain auronidin pigmentation, or have reduced amounts of both compounds. These have been used for physiological studies with respect to UVB tolerance (Clayton et al., 2018) and pathogen attack (Carella et al., 2019). In angiosperms, a range of flavonoids are localized to the cell wall (Agati et al., 2012), including rare examples of cell wall-bound anthocyanins (Philpott et al., 2009), although the physiological roles of these are generally unclear. The cell wall localization of other phenylpropanoids, in particular HCA derivatives, is common in angiosperms. These may contribute to lignin formation or be accumulated as monomers or dimers in the wall. Besides having structural roles these polymers may also contribute to physical barriers to pathogens (Zhao and Dixon, 2014). Although lignin is thought to be absent from non-vascular plants, cinnamic acid derivatives such as rosmarinic acid and (neo)lignans are common in bryophytes (Asakawa et al., 2013; Asakawa, 2017), and may be cell wall-localized (Wang et al., 2013). In *Sphagnum* moss, oxidative derivatives of sphagnum acid, *p*-hydroxyacetophenone, hydroxybutenolide, and *p*-hydroxybenzoic acid, as well as the phenolics *p*-coumaric acid and *trans*-cinnamic acid, were predominantly bound to the cell wall (Verhoeven and Liefveld, 1997). It seems a strong possibility that the red flavonoid pigments of bryophytes contribute, along with the cinnamic acid derivatives, to forming a physical barrier against pathogens. The recent study of Carella et al. (2019) demonstrated that the production of auronidin in *Marchantia*

greatly enhanced resistance to *Phytophthora palmivora* infection, with a lack of hyphae penetration into the highly pigmented regions of plants. In relation to the mechanism of action, it would be of much interest to determine the nature of the incorporation of auronidin and sphagnorubins into the wall and whether polymerization occurs. Dimers of auronidin/riccionidin A have been isolated (termed riccionidin B) (Kunz et al., 1993), providing a basis for polymerization.

Several thalloid liverwort genera, and many moss genera, have species with considerable drought tolerance, with examples in both plant groups of individuals withstanding continuous desiccation for more than 20 years (Breuil-Sée, 1993; Stark et al., 2017). As the plants of liverwort genera such as *Riccia* and *Targionia* dry out, the sides of the thallus roll over the dorsal surface so that it is covered by the darkly pigmented ventral scales and rhizoids [see **Figure 4** and Reeb et al. (2018) for examples]. This forms a “capsule” that can recover and renew growth even after extended periods without additional water. The function of the very strong pigmentation of the ventral scales, presumably by cell wall-bound auronidin, is not known. It may provide protection of the DNA against UVB damage during a period when DNA repair mechanisms are not active, given that auronidin accumulation is induced by UVB in some Antarctic species. Alternatively, the modification of the cell wall could prevent pathogen ingress, as demonstrated for *Marchantia* (Carella et al., 2019), or reduce water loss. Plants of desiccation-tolerant species in leafy liverwort genera such as *Herbertus* and *Cephaloziella* also often have dark red pigmentation (Vitt et al., 2012). A related but little studied example first described in the 1890s (Campbell, 1896) is the formation of “tubers” by some liverwort species, notably *Geothallus tuberosus*. *G. tuberosus* can form thickened inner regions of the thallus that are presumed to store carbohydrates. As the tubers form, the associated cells become strongly dark red pigmented with “thick walls”. The tubers can become buried in the soil and, although the surrounding plant may die, the thallus and associated meristem survive the long dry season of the Southern Californian regions to which the species is native.

The ventral scales of many thalloid liverworts, frequently strongly pigmented by auronidin, often extend around the apex of the thallus to provide a barrier layer between the meristem and the soil (**Figure 5**). Protection of the meristem from physical damage and pathogen ingress could explain this pigmentation. This suggestion could be extended to include protection against herbivory. The extended ventral scales of the aquatic form of *R. natans* also have strong auronidin-based purple pigmentation (**Figure 4**), so perhaps auronidins contribute to aquatic herbivore deterrence in these cells.

There are few studies on the biological functions of the 3-deoxyanthocyanins that are common in mosses and ferns, but there is evidence they also are involved in plant defense. Greater amounts of 3-deoxyanthocyanins in fronds of the aquatic fern *Azolla* correlated with increased feeding deterrence to snails and the tadpoles of frogs (Cohen et al., 2002a). In the same species, 3-deoxyanthocyanins also may promote the establishment of the symbiosis with the cyanobacterium *Nostoc* (Cohen et al., 2002b).



FIGURE 4 | Illustrations of some of the variations in form of the purple-pigmented ventral scales of thalloid liverworts. **Left** The aquatic form of *Ricciocarpos natans* (previously *Riccia natans*) has long ventral scales that extend into the water below the plant. **Right** *Riccia squamata* is one of the drought-adapted thalloid liverworts. As the environment dries out it curls over so that the ventral scales encase the thallus. Illustrations are from Lindenberg (1836).

This suggests there are specific biological functions of the flavonoid pigments in different bryophyte and fern species, although the induction of 3-deoxyanthocyanin production in ferns by general abiotic stresses also indicates a general light screening/antioxidant function in common with that of 3-hydroxyanthocyanins in angiosperms. Perhaps the cell wall-bound pigments of the bryophytes have evolved to have elegant multi-functionality, providing abiotic stress tolerance through antioxidant and/or light screening actions and altering physical properties of the cell for biotic stress resistance.

The Evolutionary Significance of the Occurrence of Different Flavonoid Structural Groups

The identification of different flavonoid groups across land plants has been conducted for many years, both to further understand the evolutionary significance of flavonoid distribution by chemotaxonomy and for the discovery of novel bioactives (Markham, 1988; Asakawa, 2017; Jiang et al., 2016; de Vries et al., 2017; Yonekura-Sakakibara et al., 2019). Flavones and/or flavonols are almost ubiquitous across land plants (Berim, 2016) but variations in the specific types of flavonols or flavones produced have occurred during evolution, for example resulting in the rarity of flavone O-glycosides in leafy liverworts (Markham, 1988) or of polymethoxylated flavones in gymnosperms (Berim, 2016). Overall, it is probable that their biosynthesis was acquired very early during land plant evolution as an important stress adaptation. The remarkable exception is the hornworts. Hornworts produce polyphenolics, notably rosmarinic acid and lignan-like compounds (e.g., anthocerotonic acid and megacerotonic acid) (Petersen and Simmonds, 2003; Soriano et al., 2018b) (**Figure 1**), but there is

no report of any flavonoid being found (Markham, 1988). Thus, either hornworts diverged from the last common land plant ancestor before the evolution of the flavonoid pathway, or the ability to make flavonoids was subsequently lost in this lineage. The completion of a hornwort genome sequence (Szövényi et al., 2015) and transcriptomic studies examining land plant evolution (Wickett et al., 2014; Puttick et al., 2018) may provide the data to help in resolving this question. Analysis of the transcriptomic (SRA PRJEB21674) and genomic (SRA ERR771108 and SRR1278954 for *Anthoceros agrestis* and *Anthoceros punctatus*, respectively) data currently available on GenBank can identify with confidence hornwort deduced sequences corresponding to the early steps of the phenylpropanoid and flavonoid pathway, including for PHENYLALANINE AMMONIA LYASE (PAL), CINNAMATE 4-HYDROXYLASE (C4H), 4-COUMARATE-COA LIGASE (4CL), CHS, and CHALCONE ISOMERASE-LIKE (CHIL) (**Supplementary Table 1**). However, additional analysis is required to show whether these produce functional enzymes. No clear CHALCONE ISOMERASE (CHI)-encoding sequence is present in the data, but mosses can make flavonoids without a gene corresponding to the typical *CHI* (Cheng et al., 2018). Rosmarinic acid is also found in some algae (Agregán et al., 2017), but comparison of the biosynthetic pathways between land plants and algae has not been made.

Until the recent clarification of the riccionidin structures as auronidins, anthocyanins were thought to be present in all extant lineages of land plants except hornworts. A progression in anthocyanin complexity was suggested, with liverworts producing “primitive” anthocyanidins (the non-glycosylated anthocyanin core), mosses and ferns 3-deoxyanthocyanins, seed plants the 3-hydroxyanthocyanins, and angiosperms a great range of substituted anthocyanins (including 5'-hydroxylation

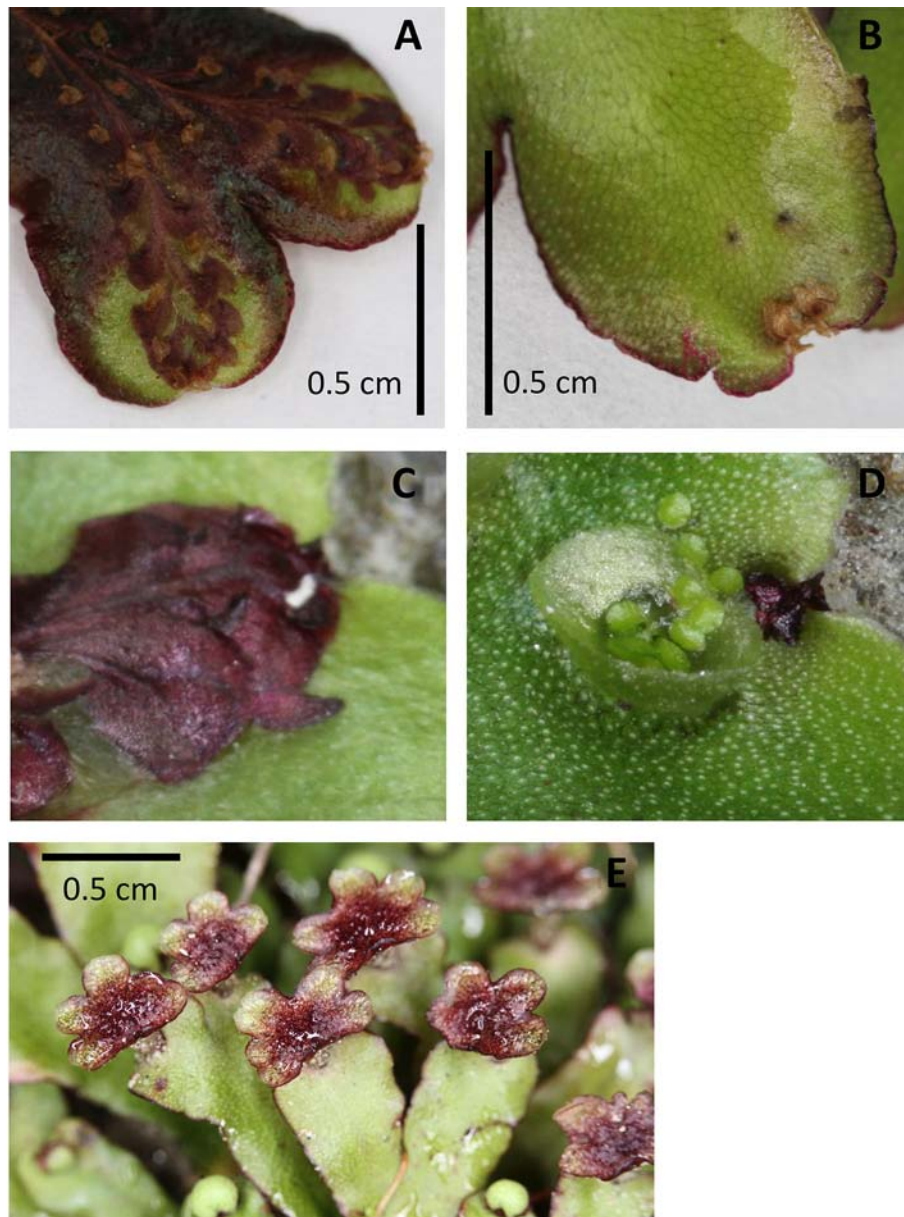


FIGURE 5 | Examples of pigmentation of the thallus and antheridiophores of *Marchantia foliacea*. Ventral (**A, C**) and dorsal (**B, D**) views of a thallus branch. The ventral surface contains the strongly pigmented scales. These extend past the meristematic notch and can be seen from the dorsal surface. The thallus in image (**A**) shows the strong pigmentation that can also occur in the non-scale cells. Image (**D**) shows a close up of the meristem region from the dorsal side, with the corresponding ventral view (**C**) “flipped” vertically to present the same orientation. Image (**E**) shows the strong pigmentation common for the surface of the antheridiophores.

and variation in glycosylation, acylation, and methylation). However, as mentioned earlier, it is now more difficult to speculate on the possible red pigments present in the last common ancestor of land plants. Riccionidin A has been reported from the root cultures of the angiosperm *Rhus chinensis* (syn. *Rhus javanica*) (Taniguchi et al., 2000), but it has not been examined whether this is synthesized *via* an aurone intermediate route.

In addition to the core flavonoid pathway found across most land plants, there are groups of flavonoids prevalent in specific taxonomic groups, such as the isoflavonoids typical of legumes. There are also flavonoid types that occur sporadically, such as aurones that are found in liverworts and some angiosperms. For aurones this may well represent convergent evolution, as even within angiosperms there are alternative biosynthetic mechanisms (Boucherle et al., 2017). New metabolomic

technologies combined with additional genome sequences for non-angiosperm species should help clarify the distribution across the land plants of different flavonoid types and the associated biosynthetic genes (Yonekura-Sakakibara et al., 2019).

The Phenylpropanoid Biosynthetic Pathway in Bryophytes

The core steps of the phenylpropanoid pathway through to the first flavonoids (the chalcones) are conserved across land plants (Tohge et al., 2013), including the presence of *PAL*, *C4H*, *4CL*, *CHS*, and *CHIL* gene sequences in hornworts. Sequences relating to some of these genes are present in the genome sequences of charophyte and chlorophyte algae (Labeeuw et al., 2015; de Vries et al., 2017), but without functional assays the conclusions that can be drawn are limited. Most phenylpropanoid pathway enzymes are thought to have evolved from primary metabolism enzymes (Tohge et al., 2013; Yonekura-Sakakibara et al., 2019), and so related sequences might be expected to be present. For *PAL*, whether it arose during land plant evolution or is an ancestral gene from algae has yet to be resolved (de Vries et al., 2017). It was suggested that *PAL* was acquired by the land plant ancestor *via* a horizontal gene transfer event (Emiliani et al., 2009), but genes related to *PAL* are present in the charophyte *Klebsormidium flaccidum* and could have been acquired by endosymbiotic gene transfer from cyanobacteria to algal ancestors of land plants (de Vries et al., 2017). *C4H*, which belongs to the CYP73A sub-family of cytochrome P450 monooxygenases (Cyp450s), shows strong sequence conservation across land plants, including characteristic motifs and residues, but no authentic gene sequences are apparent in chlorophyte genomes (Tohge et al., 2013; Davies et al., 2020). In contrast, sequences with similarity to *4CL* do occur in rhodophyte and chlorophyte genomes (Labeeuw et al., 2015; de Vries et al., 2017), suggesting the existence of this enzyme in a shared ancestor of land plants and algae before the ancestral divergence of the red algae (Labeeuw et al., 2015). A further aspect yet to be addressed is the presence in fungi of genes with significant sequence similarity to those of the phenylpropanoid pathway (Bilska et al., 2018; Lu et al., 2019). As the separation of fungi and plants is thought to have occurred during the early stages of eukaryote divergence (Burki, 2014; Burki et al., 2020), it is possible that these may represent cases of convergent evolution.

The type III PKS superfamily that contains *CHS* is present in all plant genomes examined to date (Pandith et al., 2020). PKS genes are found also in fungi, and some bacteria and algae, and the plant PKS genes contain conserved structural elements with the bacterial PKS genes involved in primary metabolism. Across plants there is a wide variety of PKS enzymes with close sequence similarity to *CHS* but which either use alternative substrates (such as acridone synthases and pyrone synthases) or catalyze different cyclisation reactions using the same starter molecules (notably, *STILBENE SYNTHASE*, *STS*). It is thought that *STS* has independently evolved from *CHS* several times in the course of evolution (Yonekura-Sakakibara et al., 2019; Pandith et al., 2020). It is probable that there are many novel PKS activities still to be discovered in plants, including bryophytes. This may include steps in bibenzyl biosynthesis, a group of liverwort

phenylpropanoid compounds related to plant defense that includes cannabinoid-like structures (Hussain et al., 2018). The presence of at least 24 PKS genes in the *Marchantia* genome suggests potential biosynthetic diversity, and at least one gene (*Mapoly0014s0122*) is closely related to the anther-specific chalcone synthase-like enzymes (ASCLs) involved in the biosynthesis of sporopollenin in angiosperms (Bowman et al., 2017). However, the majority of the annotated MpPKS genes appear to have resulted from a strongly conserved duplication of a *CHS/PAL* gene pair (Bowman et al., 2017).

The occurrence and function of *CHI* and *CHIL* in basal plants is proving to be an interesting question. Liverworts have both types of gene, and knockout *chi* mutants of *Marchantia* completely lose production of flavones (Clayton et al., 2018). Thus, in *Marchantia*, as in angiosperms examined, *CHI* is an essential *in planta* activity for flavanone production. However, no gene sequences for *CHI* have been found in moss or hornwort genome sequences or transcriptomes (Ngaki et al., 2012; Cheng et al., 2018; Berland et al., 2019). Although spontaneous closure to form the C-ring to produce flavanones from chalcones has been shown to occur *in vitro*, comparative studies on the spontaneous and enzyme catalyzed reactions suggest this is unlikely to be significant *in planta* (Jez and Noel, 2002). Studies on mutants for *chi* in *Arabidopsis* (*tt5*), carnation (*i*), and rice (*gh1*) found that flavonoid biosynthesis was not fully prevented (Stich et al., 1992; Hong et al., 2012; Jiang et al., 2015), suggesting some spontaneous conversion. However, in the case of carnation at least, the residual production of flavanones in the *chi* mutant has been found to be due to a second, weakly expressed, *CHI* gene (Miyahara et al., 2018). Thus, how flavonoid biosynthesis occurs in mosses is an open question. *CHI* and *CHIL* are thought to be examples of the rare event of catalytic activity arising in a noncatalytic scaffold protein (Kaltenbach et al., 2018). The mechanism of action of *CHIL* is unclear, and it may have differing activities across land plants, perhaps based on the promotion of activity of different biosynthetic enzymes through protein-protein interaction. In hop (*Humulus lupulus*), *HCHIL2* enhances the activities of *CHS* and an aromatic prenyltransferase (*HLPT1L*) through protein-protein interaction (Ban et al., 2018), and the promotion of flavonol and proanthocyanidin biosynthesis in *Arabidopsis* is proposed to be through direct interaction of *CHIL* and *CHI* (Jiang et al., 2015). In *Marchantia*, *CHIL* may interact with *CHS* or more than one phenylpropanoid pathway enzyme, since the production of both flavones and auronidins in *chil* mutants is only about 10% of wild-type amounts (Clayton et al., 2018). Thus, one possibility is that in mosses and hornworts *CHIL* can replace *CHI*. However, the moss *CHIL* genes assayed to date do not have *CHI* activity (Cheng et al., 2018), making this less probable.

Two major hydroxylase groups, the Cyp450s and 2-oxoglutarate dioxygenases (2OGDs, divided into the three classes DOXA, B, and C), contribute several enzymes to the phenylpropanoid pathway of angiosperms. Cyp450s include *C4H*, *FLAVONOID 3'-HYDROXYLASE* (*F3'H*), and *FLAVONE SYNTHASE II* (*FNSII*). 2OGDs include the

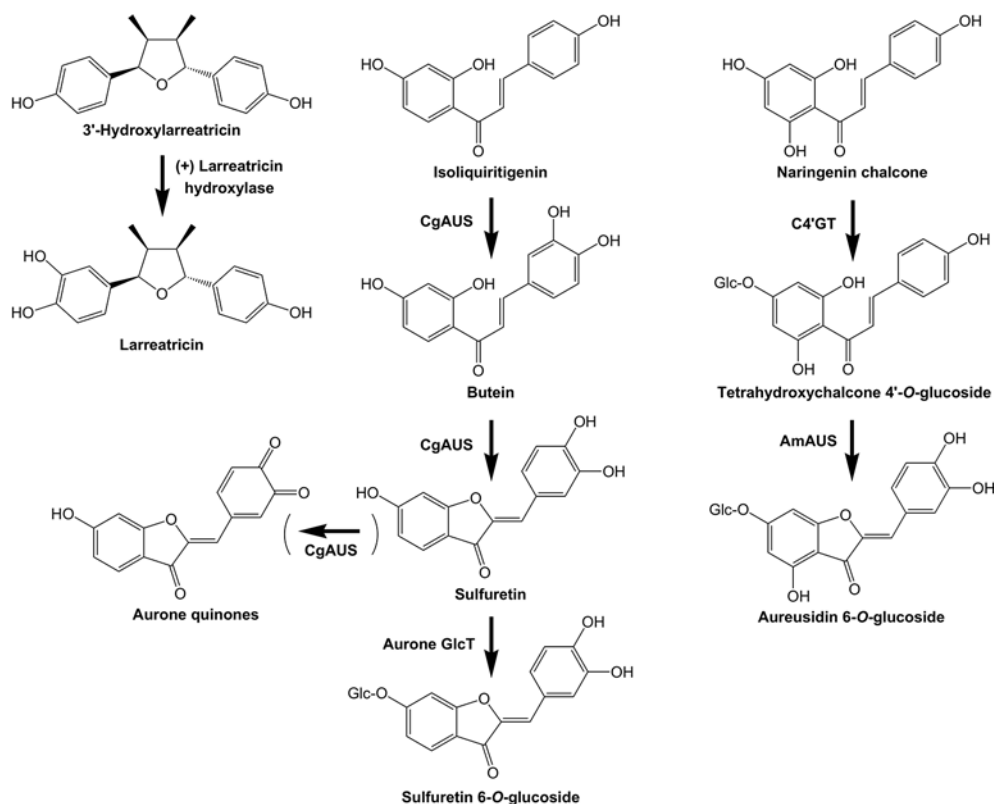


FIGURE 6 | Examples of plant POLYPHENOL OXIDASES with specific activities in specialized metabolism. The biosynthetic activities of three PPOs are shown: (+) Larreaticin hydroxylase, which is part of lignan biosynthesis in the creosote bush (*Larrea tridentata*), and the AURONE SYNTHASE from *Coreopsis grandiflora* (CgAUS) and *Antirrhinum majus* (AmAUS).

FLAVANONE 2-HYDROXYLASE (F2H), FLAVANONE 3-HYDROXYLASE (F3H), FLAVONOL SYNTHASE (FLS), FLAVONE SYNTHASE I (FNSI), and ANTHOCYANIDIN SYNTHASE/LEUCOANTHOCYANIDIN DIOXYGENASE (ANS). The evolutionary aspects of these gene families with regard to flavonoid biosynthesis were recently reviewed by Yonekura-Sakakibara et al. (2019). C4H is conserved in bryophytes, and the presence of all the other enzymes in liverworts and/or mosses would be expected based on the compounds produced. However, the close similarity of the sequences within the Cyp450 and 2OGD enzyme groups means that assignments based only on sequence similarity to the angiosperm genes should be treated with caution, and conclusive identification of other genes requires functional analysis. FNSI, F2H, and F3H have high sequence similarity and are in the DOXC28 clade and FLS and ANS are close in sequence and in the DOXC47 clade. A review of the two clades, and possible evolutionary timing of the origin of each, is given in Yonekura-Sakakibara et al. (2019).

Based on the occurrence of flavones in liverworts and mosses, it is expected that F3H and FNS activities evolved early in land plants, and two DOXC28 genes have increased transcript abundance during UVB-induced flavone production in *Marchantia* (Clayton et al., 2018). However, to date, the only

functional characterization is for a F2H that may contribute to flavone biosynthesis in the liverwort *Plagiochasma appendiculatum* (Han et al., 2014). The biosynthesis of flavones illustrates the difficulties of making assumptions about gene function, as a variety of alternative routes to flavone O- and C-glycosides have evolved in angiosperms (Jiang et al., 2016). A further complication is that the 2OGD enzymes (particularly the FLS and ANS) show promiscuous and sometimes overlapping activities when assayed *in vitro* (reviewed in Martens et al., 2010). Studies with *Arabidopsis* have shown that these “secondary” activities can also be present *in planta*, as ANS can contribute to (relatively weak) flavonol biosynthesis in the *Arabidopsis fls-1* mutant (Martens et al., 2010).

As yet, it is not clear precisely what phenylpropanoid biosynthetic activities may be present in bryophytes but not found in other plant groups. There are certainly some major pathway branches prevalent in bryophytes that are absent or rare in other groups, such as those for bibenzyls, auronidins, and sphagnorubins. Corresponding evolutionary divergence of specialized metabolic pathways would be expected to underpin the occurrence of the differing compound types. Phylogenetic analysis of the 148 Cyp450, 38 2OGD, and 41 Family-1 UDP-glycosyltransferase (the UGT family containing the “plant secondary product glycosyltransferase” motif) genes of

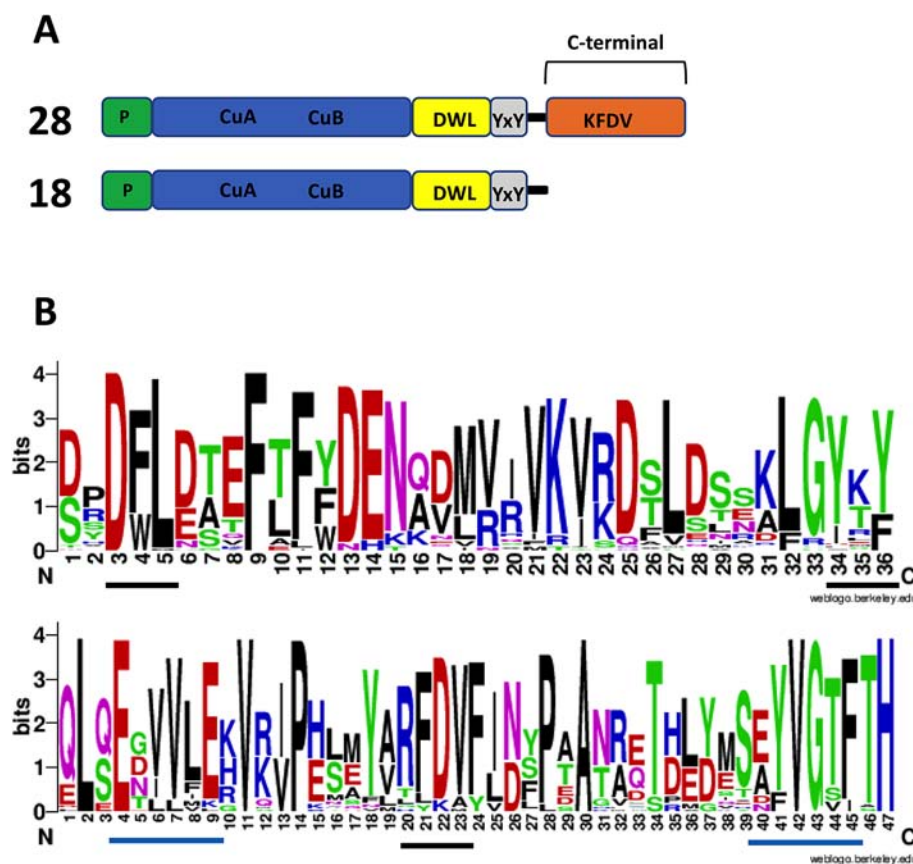


FIGURE 7 | Amino acid sequence features of polyphenol oxidases (PPOs) in *Marchantia polymorpha*. **(A)** Number of expressed PPO gene models found and their PPO type. RNA-seq data from Berland et al. (2019) were used to check for expression. The top structure (not to scale) is standard for plant PPOs. Transit/signal peptide (P), copper binding domain TYR : PFAM 00264 (CuA/CuB), PPO1_DWL: PFAM 12142 domain (DWL), tyrosine motif (YxY), PPO1_KFDV : PFAM12143 domain (KFDV). **(B)** Weblogo display of conserved residues of DWL and KFDV domains based on all gene models having an intact region (generated using <https://weblogo.berkeley.edu/logo.cgi>). DWL, tyrosine, and KFDV core motifs are underlined in black. Residues for the regions identified by Tran et al. (2012) with core motifs of EEEVLV (left) and EFAGSF (right) are underlined in blue.

Marchantia found that the majority formed individual clades that also suggested substantial lineage-specific diversification of specialized metabolism (Bowman et al., 2017). Moreover, the emerging transcriptome and genome sequence information from bryophytes is suggesting expanded functionality may have occurred for other classes of enzymes involved in phenylpropanoid biosynthesis. In the next section we examine two specialized metabolism gene families that show unexpectedly large gene family sizes in the *Marchantia* genome: those for POLYPHENOL OXIDASE (PPO) and DIRIGENT (DIR) proteins.

Liverworts May Have Expanded Functional Roles in Specialized Metabolism for Polyphenol Oxidase and Dirigent Proteins

PPO genes are found throughout land plants, as well as in bacteria, fungi, and animals, but are absent from algae. PPOs are type-III-copper proteins and the name PPO covers two major enzyme types: tyrosinases, which hydroxylate *para*-substituted monophenols to *ortho*-diphenols (monophenolase activity) and

use molecular oxygen to oxidize *ortho*-diphenols to *ortho*-quinones (diphenolase activity); and the catechol oxidases, which have only the diphenolase activity. However, it has been recently proposed that monophenolase activity could be a widespread feature of PPOs, but that the activity has remained cryptic because activity assays usually use tyrosine rather than the natural substrates, which are often not known (Molitor et al., 2016). PPOs are commonly thought of as plant defense enzymes that oxidize and/or polymerize a range of phenolic substrates with which they come into contact during cell disruption, resulting in the familiar browning reactions following tissue damage, for example in cut apples or potatoes. However, in addition to these general activities, some PPOs can conduct cross-linking reactions in biosynthetic pathways, such as latex formation; and new specific roles for PPOs have emerged in recent years (Figure 6). The published PPO gene family size in plants varies from zero (e.g., *Arabidopsis*) to 13 in *Physcomitrella patens* (Tran et al., 2012). Several angiosperm species examined have only a single PPO gene, but 11 genes have been found in

genome sequences of *Glycine max* (the legume soybean), *Populus trichocarpa* (poplar), and *Selaginella moellendorffii* (a lycophyte) (Tran et al., 2012). However, we found a much larger PPO gene family in the *Marchantia* genome: there are 64 candidate PPO genes (including gene fragments and unresolved gene models). Excluding those having partial gene models, 46 of the 64 PPO genes were represented in the RNA-seq data of Berland et al. (2019) and so are actively transcribed. Given the relatively small number of total gene models in the draft genome sequence of *Marchantia*, this represents a significant gene family, larger than the annotated 2OGD and UGT families.

Plant PPOs characterized to date are produced in a latent state as proteins of about 64–68 kDa. Besides the N-terminal targeting peptide (usually for plastid localization), PPOs contain a catalytically active domain of about 40 kDa, and a C-terminal domain of about 19 kDa that shields the active site and is later cleaved off to release the active protein. The C-terminal domain is ubiquitous in plant PPOs examined to date. Based on predicted amino acid sequences, PPOs with this typical structure are found in *Marchantia*; however, there are also members of the PPO family that lack this C-terminal domain (Figure 7), including the auronidin-related *Mapoly0021s0041*. These “short” type PPOs have also been found in fungi and bacteria (Huber et al., 1985; Shuster and Fishman, 2009; Gasparetti et al., 2010). Only a few of this short type, from the bacteria *Streptomyces* and *Bacillus*, have been extensively studied. The *Streptomyces* PPO is thought to be initially in an inactive form that is bound with a “caddie” protein. The caddie protein subsequently transfers copper to the PPO and disassociates to release an active PPO (Chen et al., 1992; Matoba et al., 2006). In contrast, the PPO from *Bacillus* does not need a caddie protein (Sendowski et al., 2011).

The first PPO found to have an unexpected role in plant specialized metabolism was the AUREUSIDIN SYNTHASE (AUS) that converts chalcone 4'-O-glucosides to aurone 6-O-glucosides in *Antirrhinum majus* (Nakayama et al., 2000; Davies et al., 2006; Ono et al., 2006; Elumalai and Liu, 2011). AmaUS differed from previously characterized PPOs in three important aspects: it was vacuole localized (Ono et al., 2006), it was a glycoprotein, and it lacked activity against common PPO substrates such as tyrosine or 3,4-dihydroxyl L-phenylalanine (L-DOPA). AmaUS conducts oxygenation of the B-ring of the chalcone, which is followed by cyclization into the aurone (Nakayama et al., 2000) (Figure 6). Although it can use chalcone aglycones *in vitro*, *in planta* aurone production in *A. majus* requires the activity of the CHALCONE 4'-O-GLUCOSYLTRANSFERASE (C4'GT) since only the glucoside is transported into the vacuole (Ono et al., 2006; Bradley et al., 2017). Subsequently, PPOs that form aurones were identified in other species, with the AURONE SYNTHASE of *Coreopsis grandiflora* that makes 4-deoxyaurones being studied in detail (Molitor et al., 2015; Molitor et al., 2016). In contrast to AmaUS, the CgAUS has the N-terminal chloroplast transit peptide and thylakoid transfer domain characteristic of plastid-localized PPOs involved in browning reactions, and uses chalcone aglycones to make aurone aglycones (Kaintz et al., 2014; Molitor et al., 2015). A PPO (*Mapoly0021s0041*) is strongly up-

regulated by MpMYB14 in association with auronidin production in *Marchantia* (Berland et al., 2019). Loss-of-function *mapoly0021s0041* mutants have greatly reduced amounts of auronidin, suggesting it too may encode an aurone biosynthetic activity, or is involved in later steps of auronidin biosynthesis and/or polymerization. Aurones have been found across land plant groups, but with sporadic occurrence. This suggests that their biosynthesis may have arisen independently on a number of occasions. Although both the aurone biosynthetic enzymes characterized to date are PPOs, the sequences are phylogenetically distinct and have differing activities and sub-cellular localization. Additionally, the biosynthesis of the aurone hispidol in *Medicago truncatula* may be conducted by a peroxidase rather than a PPO (Farg et al., 2009). Besides AUS, PPOs have been implicated in tyrosine or phenylpropanoid biosynthetic pathways of walnut (Araji et al., 2014) and creosote bush (*Larrea tridentata*) (Cho et al., 2003).

Forty of the *Marchantia* PPO genes (66%) occur as tandem repeats or small gene clusters (local tandemly arrayed genes; TAGs). Although this figure may be either an under- or over-estimate as it is based on the initial scaffold assembly of the genome (Bowman et al., 2017), it is nevertheless a much higher value than the overall percentage of *Marchantia* TAGs estimated on the same basis, which at 5.9% is near the lower end of the range observed in flowering plants (Bowman et al., 2017). The TAG percentage is also relatively high for some of the other characterized specialized metabolite gene families of *Marchantia*. For example, there are 18 occurrences of neighboring PAL and/or CHS genes. TAGs are notable in some angiosperm species that have prominent specialized metabolic characteristics—such as the terpenoid pathways of the tree species *Eucalyptus grandis* and teak (*Tectona grandis*). Teak has at least 14 TAGs for the terpene synthase gene family (Zhao et al., 2019). *E. grandis* has the largest number of genes in tandem repeats reported among sequenced plant genomes, at 34% of total genes (Myburg et al., 2014). For the *Marchantia* phenylpropanoid biosynthetic pathway, 10 multigene families have expanded, mostly through tandem duplication, to result in a total of 174 genes. In angiosperms, gene diversification is a result of a combination of local duplication events and whole genome duplications, but it is probable that no whole-genome duplication events have occurred during liverwort evolution (Bowman et al., 2017). Therefore, although there is no overall increase in the frequency of TAGs in liverworts (at least for *Marchantia*), local gene duplication events are likely to have been a common mechanism for generating gene neofunctionalization in specialized metabolism. Whether this is typical of other liverworts requires the completion of further genome sequences. However, BLAST analysis of the *Lunularia cruciata* transcriptome (www.polebio.lrsv.ups-tlse.fr/Luc_v1/Luc_v1.fa) identified more than 20 sequences with the conserved features of PPOs (data not shown), suggesting a large gene family in this species also.

Dirigent proteins (DIR) are small (~16–18 kDa) cell wall-localized proteins that may control the regio- and stereospecific outcome of phenoxy radical coupling in lignin and lignan

polymerization reactions (Davin et al., 1997; Gang et al., 1999). The polymerization reactions also require the activity of laccase or peroxidase to produce electron oxidative capacity to generate the phenoxy radical. In vascular plants, lignins are complex, amorphous heteropolymers involved in wall strengthening and pathogen resistance, with species-specific composition produced by polymerization of coniferyl, sinapyl, and *p*-coumaryl alcohols. A role for DIRs in directing the reactions leading to the formation of lignin has been proposed but not definitively established, although there is strong genetic evidence in support of some specific cases (Hosmani et al., 2013).

The role of DIRs in determining stereospecificity has been best described in the formation of lignans, a class of 8-8' linked C₆C₃ phenylpropanoid dimers involved in pathogen resistance, for example in the production of (+)- or (–)-pinoresinol compounds in flax (*Linum usitatissimum*) and pea (*Pisum sativum*) (Kim et al., 2015; Corbin et al., 2018). The X-ray crystal structure of PsDRR206 involved in (+)-pinoresinol formation suggested that the active protein had a trimeric structure (Kim et al., 2015). Recent work has suggested that at least some DIRs may do more than the hypothesized positioning of phenoxy radicals prior to coupling, and may themselves possess enzymatic activity. The crystal structure of *Arabidopsis* AtDIR6 identified potentially catalytic residues including aspartic acids that were essential for activity, and it was proposed that this protein catalyzed the cyclization of the bis-quinone methide intermediate during (+)- or (–)-pinoresinol formation (Gasper et al., 2016). Also, a recombinantly expressed DIR from *Glycyrrhiza echinata* was found to possess isoflavanol dehydratase activity and carry out the final ring-closure step of the biosynthesis of the anti-microbial phytoalexin (–)-pterocarpan (Uchida et al., 2017).

DIR gene families can be quite large, with 26 genes in *Arabidopsis* (Paniagua et al., 2017) and 44 genes in flax (of which seven appeared to be gene fragments or result from chromosomal rearrangements; Corbin et al., 2018). Of the 37 genes with classical DIR structure in flax, 15 paralogous gene pairs were identified. Kubo et al. (2018) identified 52 dirigent-like predicted proteins in the *Marchantia* genome sequence. Our analysis for this article found that at least 35 of these occur as TAGs. However, the deduced protein sequences of the family members are diverse, and the functionality of the proteins has not yet been established. Our BLAST analysis of the *Marchantia* genome and transcript resources with the 24 annotated *Arabidopsis* DIR genes gave us 60 initial candidate gene models, with strong evidence of some very recent gene duplications giving groups of adjacent genes with highly similar or identical deduced amino acid sequences.

Phenylpropanoid biosynthesis and lignification are common plant responses to biotic and abiotic stress (Zhao and Dixon, 2014; Paniagua et al., 2017) and consequently, as a component of lignification, DIRs have been implicated in responses to pathogen and drought stress (e.g., Thamil Arasan et al., 2013; Paniagua et al., 2017). In *P. patens*, fungal infection resulted in increased incorporation of phenolic compounds into the wall and up-regulation of a DIR gene (Reboledo et al., 2015). In

Marchantia, abiotic stresses such as UVB irradiation, N deficiency and salinity (Albert et al., 2018; Kubo et al., 2018), and pathogen attack (Carella et al., 2019) increased the expression of MpMYB14. MpMYB14 promotes auronidin production and up-regulates transcript abundance for at least three DIR genes (*Mapoly0006s0216*, *Mapoly0006s0217*, *Mapoly0078s0058*) (Albert et al., 2018; Kubo et al., 2018). The deduced protein products of these genes possess predicted signal peptides (SignalP 5.0, <http://www.cbs.dtu.dk/services/SignalP/>), indicative of secretion to the vacuole or, extracellularly, to the cell wall. However, prediction of subcellular localization using WoLFPSORT (<https://wolfpsort.hgc.jp>) indicated with low confidence different compartments for the three predicted protein products of the genes: extracellular (*Mapoly0006s0217*), vacuolar (*Mapoly0006s0216*), and cytoplasmic (*Mapoly0078s0058*). Intracellular coupling of monolignol radicals has been described in *Arabidopsis* (Dima et al., 2015). There have been no studies on DIR genes in hornworts. However, as lignans are prominent specialized metabolites of hornworts, and DIRs have roles in lignan biosynthesis in angiosperms, this could be a worthwhile area to investigate.

Evolution of the Transcriptional Regulation of the Phenylpropanoid Pathway

In angiosperms and gymnosperms, the key regulatory complex consists of R2R3MYB and bHLH TFs joined with a WD-Repeat (WDR) protein, a composition of proteins known as an “MYB-bHLH-WD repeat (MBW)” complex. The MBW complex that activates anthocyanin and proanthocyanidin production contains R2R3MYB proteins from sub-group (SG) 5 or 6, and commonly promotes transcription of the biosynthetic genes throughout the pathway. The action of the MBW complex is modified by a WRKY class activator TF and a series of proteins with repressor actions (Lloyd et al., 2017). In particular, R2R3MYBs from SG4 can join an activating MBW complex and turn it into one that represses target gene transcription, and R3MYBs can bind the bHLH to prevent it from forming the MBW complex, thus competitively inhibiting activation (Albert et al., 2014). The SG4 R2R3MYBs are characterized by the presence of an ethylene response factor (ERF)-associated amphiphilic repression (EAR) motif (LxLxL or DLNxxP) or a TLLFR motif in the C terminus that mediates transcriptional repression (Chen et al., 2019a; Chen et al., 2019b; Ma and Constabel, 2019). For the activation of the flavonol and flavone branches, a SG7 R2R3MYB acts without being part of the complex. There is also regulation upstream of the flavonoid pathway, as HY5 activates the production of the SG7 R2R3MYB. Additionally, in *Arabidopsis* it has been shown that HY5 directly activates transcription of some flavonoid biosynthetic genes, such as CHS. The conservation of HY5 function in the UVB responses of both bryophytes and angiosperms was mentioned earlier, although its target gene set has yet to be resolved.

The expansion of TF families during evolution has been a driver of diversity in land plants, as a consequence of multicellularity and increased organismal complexity and/or for coping with the increased stress of a sessile land-based

lifestyle. The MYB gene family is one of the largest TF families in plants, with *Arabidopsis* having 137 R2R3MYB genes (Feller et al., 2011). This includes one SG5, four SG6, and three SG7 genes in *Arabidopsis* for proanthocyanidin, anthocyanin, and flavonol production, respectively. The presence of small gene families for sub-groups regulating specialized metabolic pathways is common for angiosperms, and has enabled sub-functionalization and diversification of flavonoid temporal and spatial regulation in flowers, seeds, and vegetative tissues. The bHLH and WDR components are less specific in their regulatory targets, and can regulate other characters as well as flavonoid biosynthesis, such as epidermal cell differentiation.

The great majority of information on the transcriptional regulation of specialized metabolite pathways is available from studies on angiosperms, with only a small number of studies on gymnosperm, fern, or bryophyte species. Identifying the genetic components for flavonoid pathway regulation in these other plant groups will help establish a model for how regulation of specialized metabolism may have changed during evolution. For bryophytes, notable questions relating to flavonoid pathway regulation include: are R2R3MYB and bHLH genes the key direct activators? If MYBs are the direct activators, which SGs are present in bryophytes and do small gene families occur for each SG? Does a MBW complex form in bryophytes? Do repressor TFs modify pathway regulation? Characterizing these aspects in species such as *M. polymorpha*, *P. patens*, and the lycophyte model *S. moellendorffii* should indicate which aspects of flavonoid regulation are conserved across land plants, and thus may have been present in the early land plant ancestor, and which aspects may have arisen as part of evolutionary diversification of the different land plant groups.

Compared with angiosperms, the characterized bryophytes and lycophytes have small TF families. There are only 22, 49, and 62 R2R3MYBs in the genomes of *M. polymorpha*, *P. patens*, and *S. moellendorffii*, respectively (Feller et al., 2011; Bowman et al., 2017). For Marchantia genes, a phylogenetic comparison of this gene family shows that *MpMYB02* and *MpMYB14* fall basal to a clade that contains all the phenylpropanoid-related R2R3MYB genes of *Arabidopsis* (SGs 4, 5, 6, 7, 15, and 44) (Bowman et al., 2017). Concluding whether these correspond to descendants of the flavonoid regulatory R2R3MYBs of the ancestral land plant requires further study, although both *MpMYB02* and *MpMYB14* activate phenylpropanoid biosynthetic genes. *MpMYB02* is required for production of bibenzyls (Kubo et al., 2018) while *MpMYB14* is essential for auronidin production and promotes the production of flavone O-glycosides (Albert et al., 2018; Clayton et al., 2018; Kubo et al., 2018). The profiles of transcripts up-regulated by *MpMYB02* and *MpMYB14* include *DIR* genes (Albert et al., 2018; Kubo et al., 2018; Berland et al., 2019). For *MpMYB14*, this includes the three *DIR* genes discussed earlier as well as other *DIR* genes that have been shown to be direct targets (Kubo et al., 2018). Co-expression analysis in flax found that MYB TFs were up-regulated along with *DIR* genes during secondary wall biosynthesis (Corbin et al., 2018), suggesting that MYB proteins could control *DIR* expression in both angiosperms and bryophytes.

MpMYB14 must act redundantly with other uncharacterized TFs for flavone production, as *Mpmyb14* mutants still show the induction of flavones in response to UVB (Clayton et al., 2018), nutrient stress, or high-irradiance white light (Albert et al., 2018). Flavone production is reduced in *Mphy5* mutants, so it is possible that HY5 is a direct activator of flavonoid biosynthetic genes as in *Arabidopsis*, but there may also be HY5-independent activation pathways for flavone production (Kondou et al., 2019). Analysis of changes in transcriptomes in response to UVB treatment does not present any alternative R2R3MYB candidate for flavone regulation (Clayton et al., 2018). Thus, Marchantia may lack the equivalent of the angiosperm SG7 activators of flavonol and flavone biosynthesis.

The *PabHLH* gene of the liverwort *P. appendiculatum* is a probable activator of bibenzyl biosynthesis (Wu et al., 2018). Over-expression of *PabHLH* in *P. appendiculatum* increased bibenzyl concentration and up-regulated transcript abundance from known phenylpropanoid biosynthetic genes (*PAL*, *4CL*) and candidate bibenzyl biosynthetic genes, whereas RNA interference-induced suppression down-regulated the same genes and reduced bibenzyl accumulation. Phylogenetically, *PabHLH* falls within clades containing the flavonoid MBW bHLH sequences of angiosperms (within bHLH subgroup IIIIf), suggesting it may be homologous to them. In Marchantia, *MpbHLH12* is the gene with the highest sequence identity to *PabHLH* and the flavonoid-related bHLHs of angiosperms, and transcriptomic analysis of *MpbHLH12* overexpression transgenics suggests it may also be involved in flavonoid regulation (Arai et al., 2019). However, although R2R3MYB and bHLH genes do regulate flavonoid biosynthesis in liverworts, and there are conserved WDR sequences in the genome (Bowman et al., 2017), there is no answer yet on whether the MBW complex exists in bryophytes. A flavonoid-related MBW complex has been characterized in the gymnosperm Norway spruce (*Picea abies*) (Nemesio-Gorritz et al., 2017), supporting an origin for the MBW complex in the plant lineage prior to the last common ancestor of gymnosperms and angiosperms, around 350–300 MYA. However, although the conserved amino acid motif ([D/E]Lx2[R/K]x3Lx6Lx3R) identified as necessary for R2R3MYB proteins to bind the bHLH partners (Zimmermann et al., 2004) is present in the *S. moellendorffii* sequence SmXP002978781, it is lacking in bryophyte R2R3MYBs studied to date. The closest matches in *P. patens* (PpXP001752936) and Marchantia (*MpMYB02* and *MpMYB14*) lack one and two deduced amino acid residues, respectively.

Whether bryophytes possess MYB genes with a repressive action in phenylpropanoid regulation, either the R2R3MYB active repressors that form part of the MBW complex or the R3MYBs that are thought to “compete” for the bHLH proteins, is also an open question. We were unable to identify (known) repression motifs in any of the Marchantia R2R3MYB sequences. Both *P. patens* and *S. moellendorffii* have R2R3MYB genes with putative EAR motif sequences (LxLxL), but the possible function of these in regulating phenylpropanoid biosynthesis has not been examined. Analysis of the auxin signaling pathway of

Marchantia has identified an orthologue of TOPLESS, which in angiosperms interacts with the EAR motif to mediate transcriptional repression (Flores-Sandoval et al., 2015). The Marchantia genome contains an expanded R3MYB gene family (Bowman et al., 2017), but no analysis of these with regard to flavonoid biosynthesis has been published.

In summary, based on the evidence from Marchantia, it seems probable that the ancestral R2R3MYB regulators of phenylpropanoid metabolism were activators acting outside of an MBW complex. R2R3MYB-repressive TFs and the MBW complex probably evolved after the last common ancestor of liverworts and gymnosperms/angiosperms. As the flavone pathway probably evolved prior to anthocyanin biosynthesis, it could be expected that R2R3MYBs most similar to SG7 might be the ancestral type. However, the specific flavone activators of Marchantia have yet to be identified. MpMYB02 and MpMYB14 may correspond to the ancestral phenylpropanoid pathway activators, and like SG7 probably act outside the MBW complex, but it is difficult to state which is the most closely related SG because of the extent of sequence divergence, with no conservation of sequence outside the MYB domains themselves. Furthermore, additional data are required from other bryophyte species, as the evolutionary path to Marchantia will have resulted in extensive genetic changes and the loss of characters that were present in the last common ancestor.

CONCLUDING COMMENTS

The commonality of phenylpropanoid biosynthetic genes between bryophytes and angiosperms, and the conserved functions of flavonoids in assisting in tolerance to stresses such as UVB and pathogen attack, support the proposal that the pathway arose before the last common ancestor of these land plant groups, relatively early during the process of land colonization. The exception to this is the hornworts, which lack flavonoids. Unless the divergence of hornworts occurred before the pathway arose, the hornwort ancestor must have acquired mutations that caused loss of the biosynthetic or regulatory capacity. This may be analogous to the loss of anthocyanin biosynthesis in some lineages of the Caryophyllales, where they are replaced by betalains. As the main red pigments of angiosperms (soluble anthocyanins) and bryophytes (cell wall-bound auronidins and sphagnorubins) differ in structure and cellular properties, it is difficult to suggest what the original common ancestor may have possessed with regard to red pigments. Establishing which components of anthocyanin biosynthesis are present or lacking in bryophytes may help in this regard.

REFERENCES

Agati, G., and Tattini, M. (2010). Multiple functional roles of flavonoids in photoprotection. *New Phytol.* 186, 786–793. doi: 10.1111/j.1469-8137.2010.03269.x

The diversification of both specialized metabolite biosynthesis and the transcription factors that regulate the pathways are thought to be important contributors to the evolution of plants to occupy the varied ecological niches offered on land (Pichersky and Gang, 2000). To date, much of our understanding of the genetic basis of the diversification process has been based on studies of flowering plants. However, the completion of the first genome sequences for a moss (*P. patens*) and liverwort (*M. polymorpha*) has started to reveal the details of the specialized metabolite gene families, such as for phenylpropanoid biosynthesis. Notably, the TF families thought to regulate the phenylpropanoid pathway are much smaller in Marchantia than in flowering plants. However, there are relatively large Marchantia gene families for enzymes that are often involved in specialized metabolism, such as the Cyp450, 2OGD, and UGT families. Moreover, Marchantia has large PPO and DIR gene families compared to angiosperms, suggesting these enzyme groups may make a greater contribution than previously anticipated to phenylpropanoid and other specialized metabolite biosynthesis in the liverworts. Thus, in liverworts some of the gene families involved in the biosynthesis of specialized metabolites appear to have undergone more gene duplication (allowing consequent sub- and neofunctionalization for particular family members) than the TFs that regulate the same pathways. Expansion of the regulatory TF families through duplication and sub/neofunctionalization is seen in the angiosperms, probably reflecting increased organismal complexity.

AUTHOR CONTRIBUTIONS

KD, RJ, YZ, NA, DB, BJ, JB and KS reviewed literature, formulated ideas and wrote the manuscript. KD and KS prepared the figures. KD, RJ, YZ, DB and KS conducted bioinformatic analysis.

FUNDING

Financial support was provided by the Marsden Fund of New Zealand Grant PAF1701.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2020.00007/full#supplementary-material>

Agati, G., Azzarello, E., Pollastri, S., and Tattini, M. (2012). Flavonoids as antioxidants in plants: Location and functional significance. *Plant Sci.* 196, 67–76. doi: 10.1016/j.plantsci.2012.07.014

Agregán, R., Munekata, P. E. S., Franco, D., Dominguez, R., Carballo, J., and Lorenzo, J. M. (2017). Phenolic compounds from three brown seaweed species

- using LC-DAD-ESI-MS/MS. *Food Res. Int.* 99, 979–985. doi: 10.1016/j.foodres.2017.03.043
- Aigner, S., Remias, D., Karsten, U., and Holzinger, A. (2013). Unusual phenolic compounds contribute to ecophysiological performance in the purple-colored green alga *Zygonium ericetorum* (zygnematophyceae, streptophyta) from a high-alpine habitat. *J. Phycol.* 49, 648–660. doi: 10.1111/jpy.12075
- Albert, N. W., Davies, K. M., Lewis, D. H., Zhang, H., Montefiori, M., Brendolise, C., et al. (2014). A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in eudicots. *Plant Cell* 26, 962–980. doi: 10.1105/tpc.113.122069
- Albert, N. W., Thirumawithana, A. H., Mcghee, T. K., Clayton, W., Deroles, S. C., Schwinn, K. E., et al. (2018). Genetic analysis of the liverwort *Marchantia polymorpha* reveals that R2R3MYB activation of flavonoid production in response to abiotic stress is an ancient character in land plants. *New Phytol.* 218, 554–566. doi: 10.1111/nph.15002
- Allorent, G., Lefebvre-Legendre, L., Chappuis, R., Kuntz, M., Truong, T. B., Niyogi, K. K., et al. (2016). UV-B photoreceptor-mediated protection of the photosynthetic machinery in *Chlamydomonas reinhardtii*. *Proc. Natl. Acad. Sci. U. S. A.* 113, 14864–14869. doi: 10.1073/pnas.1607695114
- Andersen O. M. and Markham K. R. (Eds.) (2006). *The Flavonoids: Chemistry, Biochemistry and Applications*. (Boca Raton, USA: CRC Press).
- Arai, H., Yanagiura, K., Toyama, Y., and Morohashi, K. (2019). Genome-wide analysis of MpBHLH12, a III/II basic helix-loop-helix transcription factor of *Marchantia polymorpha*. *J. Plant Res.* 132, 197–209. doi: 10.1007/s10265-019-01095-w
- Araji, S., Grammer, T. A., Gertzen, R., Anderson, S. D., Mikulic-Petkovsek, M., Veberic, R., et al. (2014). Novel roles for the polyphenol oxidase enzyme in secondary metabolism and the regulation of cell death in walnut. *Plant Physiol.* 164, 1191–1203. doi: 10.1104/pp.113.228593
- Archetti, M. (2009). Classification of hypotheses for the evolution of autumn colours. *Oikos* 118, 328–333. doi: 10.1111/j.1600-0706.2008.17164.x
- Asakawa, Y., Ludwiczuk, A., and Nagashima, F. (2013). “Chemical Constituents of Bryophytes: Bio- and Chemical Diversity, Biological Activity, and Chemosystematics,” in *Progress in the Chemistry of Organic Natural Products*. Eds. A. D. Kinghorn, H. Falk and J. Kobayashi (Vienna: Springer), 1–796. doi: 10.1007/978-3-7091-1084-3
- Asakawa, Y. (2017). “Polyphenols in Bryophytes,” in *Recent Advances in Polyphenol Research*. Eds. K. Yoshida, V. Cheynier and S. Quideau, Wiley-Blackwell: Oxford, UK. 36–66. doi: 10.1002/9781118883303.ch2
- Ban, Z., Qin, H., Mitchell, A. J., Liu, B., Zhang, F., Weng, J.-K., et al. (2018). Noncatalytic chalcone isomerase-fold proteins in *Humulus lupulus* are auxiliary components in prenylated flavonoid biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* 115, E5223–E5232. doi: 10.1073/pnas.1802231115
- Ben Saad, H., Gargouri, M., Kallel, F., Chaabene, R., Boudawara, T., Jamoussi, K., et al. (2017). Flavonoid compounds from the red marine alga *Alsidium corallinum* protect against potassium bromate-induced nephrotoxicity in adult mice. *Environ. Toxicol.* 32, 1475–1486. doi: 10.1002/tox.22368
- Berim, A. (2016). Methoxylated flavones: occurrence, importance, biosynthesis. *Phytochem. Rev.* 15, 363–390. doi: 10.1007/s11101-015-9426-0
- Berland, H., Albert, N. W., Stavland, A., Jordheim, M., McGhee, T. M., Zhou, Y., et al. (2019). Auronidins are a previously unreported class of flavonoid pigments that challenges when anthocyanin biosynthesis evolved in plants. *Proc. Natl. Acad. Sci. U. S. A.* 116, 20232–20239. doi: 10.1073/pnas.1912741116
- Bilka, K., Stuper-Szablewska, K., Kulik, T., Buško, M., Załuski, D., Jurczak, S., et al. (2018). Changes in phenylpropanoid and trichothecene production by *Fusarium culmorum* and *F. graminearum* Sensu Stricto via exposure to flavonoids. *Toxins* 10, 110. doi: 10.3390/toxins10030110
- Bonnett, S. A. F., Ostle, N., and Freeman, C. (2010). Short-term effect of deep shade and enhanced nitrogen supply on *Sphagnum capillifolium* morphophysiology. *Plant Ecol.* 207, 347–358. doi: 10.1007/s11258-009-9678-0
- Boucherle, B., Peuchmaur, M., Boumendjel, A., and Haudecoeur, R. (2017). Occurrences, biosynthesis and properties of auronides as high-end evolutionary products. *Phytochemistry* 142, 92–111. doi: 10.1016/j.phytochem.2017.06.017
- Bowman, J. L., Araki, T., and Kohchi, T. (2016). *Marchantia*: Past, Present and Future. *Plant Cell Physiol.* 57, 205–209. doi: 10.1093/pcp/pcw023
- Bowman, J. L., Kohchi, T., Yamato, K. T., Jenkins, J., Shu, S., Ishizaki, K., et al. (2017). Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. *Cell* 171, 287–304. doi: 10.1016/j.cell.2017.09.030
- Bradley, D., Xu, P., Mohorianu, I.-I., Whibley, A., Field, D., Tavares, H., et al. (2017). Evolution of flower color pattern through selection on regulatory small RNAs. *Science* 358, 925–928. doi: 10.1126/science.aao3526
- Breuil-Sée, A. (1993). Recorded desiccation-survival times in bryophytes. *J. Bryol.* 17, 679–680. doi: 10.1179/jbr.1993.17.4.679
- Brown, D. E., Rashotte, A. M., Murphy, A. S., Normanly, J., Tague, B. W., Peer, W. A., et al. (2001). Flavonoids act as negative regulators of auxin transport *in vivo* in *Arabidopsis*. *Plant Physiol.* 126, 524–535. doi: 10.1104/pp.126.2.524
- Brunetti, C., Fini, A., Sebastiani, F., Gori, A., and Tattini, M. (2018). Modulation of phytohormone signaling: a primary function of flavonoids in plant–environment interactions. *Front. Plant Sci.* 9, 1042. doi: 10.3389/fpls.2018.01042
- Brunetti, C., Sebastiani, F., and Tattini, M. (2019). Review: ABA, flavonols, and the evolvability of land plants. *Plant Sci.* 280, 448–454. doi: 10.1016/j.plantsci.2018.12.010
- Buer, C. S., and Muday, G. K. (2004). The *transparent testa4* mutation prevents flavonoid synthesis and alters auxin transport and the responses of *Arabidopsis* roots to gravity and light. *Plant Cell* 16, 1191–1205. doi: 10.1105/tpc.020313
- Burbulis, I. E., Iacobucci, M., and Shirley, B. W. (1996). A null mutation in the first enzyme of flavonoid biosynthesis does not affect male fertility in *Arabidopsis*. *Plant Cell* 8, 1013–1025. doi: 10.1105/tpc.8.6.1013
- Burki, F., Roger, A. J., Brown, M. W., and Simpson, A. G. B. (2020). The new tree of eukaryotes. *Trends Ecol. Evol.* in press. 43–55.
- Burki, F. (2014). The eukaryotic tree of life from a global phylogenomic perspective. *Cold Spring Harb. Perspect. Biol.* 6, a016147. doi: 10.1101/cshperspect.a016147
- Campbell, D. H. (1896). The development of *Geothallus tuberosus*, Campbell. *Ann. Bot.* 10, 489–510. doi: 10.1093/oxfordjournals.aob.a088624
- Carella, P., Gogleva, A., Hoey, D. J., Bridgen, A. J., Stolze, S. C., Nakagami, H., et al. (2019). Conserved biochemical defenses underpin host responses to oomycete infection in an early divergent land plant lineage. *Curr. Biol.* 29, 1–13. doi: 10.1016/j.cub.2019.05.078
- Cesares Gil, D. A. (1902). Algunas observaciones sobre la coloracion rojiza de ciertas hepáticas. *Bol. Soc. Española Hist. Nat.* 2, 217–220.
- Chattopadhyay, A., Erland, L. A. E., Jones, A. M. P., and Saxena, P. K. (2018). Indoleamines and phenylpropanoids modify development in the bryophyte *Plagiommium cuspidatum* (Hedw.) T. J. Kopp. *In Vitro Cell. Dev. Biol. Plant* 54, 454–464. doi: 10.1007/s11627-018-9904-3
- Chen, L.-Y., Leu, W.-M., Wang, K.-T., and Lee, Y.-H. W. (1992). Copper transfer and activation of the *Streptomyces* apotyrasinase are mediated through a complex formation between apotyrasinase and its trans-activator MelC1. *J. Biol. Chem.* 267, 20100–20107.
- Chen, C., Zhang, K., Khurshid, M., Li, J., He, M., Georgiev, M. I., et al. (2019a). MYB transcription repressors regulate plant secondary metabolism. *Crit. Rev. Plant Sci.* 38, 159–170. doi: 10.1080/07352689.2019.1632542
- Chen, L., Hu, B., Qin, Y., Hu, G., and Zhao, J. (2019b). Advance of the negative regulation of anthocyanin biosynthesis by MYB transcription factors. *Plant Physiol. Biochem.* 136, 178–187. doi: 10.1016/j.plaphy.2019.01.024
- Cheng, A.-X., Zhang, X., Han, X.-J., Zhang, Y.-Y., Gao, S., Liu, C.-J., et al. (2018). Identification of chalcone isomerase in the basal land plants reveals an ancient evolution of enzymatic cyclization activity for synthesis of flavonoids. *New Phytol.* 217, 909–924. doi: 10.1111/nph.14852
- Cheyrier, V., Comte, G., Davies, K. M., Lattanzio, V., and Martens, S. (2013). Plant phenolics: Recent advances on their biosynthesis, genetics, and ecophysiology. *Plant Physiol. Biochem.* 72, 1–20. doi: 10.1016/j.plaphy.2013.05.009
- Cho, M. H., Moinuddin, S. G., Helms, G. L., Hishiyama, S., Eichinger, D., Davin, L. B., et al. (2003). (+)-Larreticrin hydroxylase, an enantio-specific polyphenol oxidase from the creosote bush (*Larrea tridentata*). *Proc. Natl. Acad. Sci. U. S. A.* 100, 10641–10646. doi: 10.1073/pnas.1934562100
- Clarke, L. J., and Robinson, S. A. (2008). Cell wall-bound ultraviolet-screening compounds explain the high ultraviolet tolerance of the Antarctic moss, *Ceratodon purpureus*. *New Phytol.* 179, 776–783. doi: 10.1111/j.1469-8137.2008.02499.x
- Clayton, W. A., Albert, N. W., Thirumawithana, A. H., McGhee, T. K., Deroles, S. C., Schwinn, K. E., et al. (2018). UVR8-mediated induction of flavonoid biosynthesis for UVB tolerance is conserved between the liverwort

- Marchantia polymorpha* and flowering plants. *Plant J.* 96, 503–517. doi: 10.1111/tpj.14044
- Cockell, M. M., and Knowland, J. (1999). Ultraviolet radiation screening compounds. *Biol. Rev.* 74, 311–345. doi: 10.1111/j.1469-185X.1999.tb00189.x
- Cohen, M. F., Meziane, T., Tsuchiya, M., and Yamasaki, H. (2002a). Feeding deterrence of *Azolla* in relation to deoxyanthocyanin and fatty acid composition. *Aquat. Bot.* 74, 181–187. doi: 10.1016/S0304-3770(02)00077-3
- Cohen, M. F., Sakihama, Y., Takagi, Y. C., Ichiba, T., and Yamasaki, H. (2002b). Synergistic effect of deoxyanthocyanins from symbiotic fern *Azolla* spp. on *hrmA* gene induction in the cyanobacterium *Nostoc punctiforme*. *Mol. Plant Microbe Interact.* 15, 875–882. doi: 10.1094/MPMI.2002.15.9.875
- Corbin, C., Drouet, S., Markulin, L., Auguin, D., Laine, E., Davin, L. B., et al. (2018). A genome-wide analysis of the flax (*Linum usitatissimum* L.) dirigent protein family: from gene identification and evolution to differential regulation. *Plant Mol. Biol.* 97, 73–101. doi: 10.1007/s11103-018-0725-x
- Dare, A. P., and Hellens, R. P. (2013). RNA interference silencing of CHS greatly alters the growth pattern of apple (*Malus x domestica*). *Plant Signal. Behav.* 8, e25033. doi: 10.4161/psb.25033
- Davies, K. M., Marshall, G. B., Bradley, J. M., Schwinn, K. E., Bloor, S. J., Winefield, C. S., et al. (2006). Characterisation of aurone biosynthesis in *Antirrhinum majus*. *Physiol. Plant* 128, 593–603. doi: 10.1111/j.1399-3054.2006.00774.x
- Davies, K., Albert, N., Zhou, Y., and Schwinn, K. (2018). Functions of flavonoid and betalain pigments in abiotic stress tolerance in plants. *Annu. Plant Rev. Online* 1, 1–41. doi: 10.1002/9781119312994.apr0131
- Davies, K. M., Jibrán, R., Albert, N. W., Zhou, Y., and Schwinn, K. E. (2020). “Conservation and divergence between bryophytes and angiosperms in the biosynthesis and regulation of flavonoid production,” in *Recent Advances in Polyphenols Research*. Eds. S. Quideau, J. Reed and V. De Freitas (Oxford, UK: Wiley-Blackwell). in press.
- Davin, L. B., Wang, H. B., Crowell, A. L., Bedgar, D. L., Martin, D. M., Sarkanen, S., et al. (1997). Stereoselective bimolecular phenoxy radical coupling by an auxiliary (dirigent) protein without an active center. *Science* 275, 362–366. doi: 10.1126/science.275.5298.362
- de Vries, J., and Archibald, J. M. (2018). Plant evolution: landmarks on the path to terrestrial life. *New Phytol.* 217, 1428–1434. doi: 10.1111/nph.14975
- de Vries, J., de Vries, S., Slamovits, C. H., Rose, L. E., and Archibald, J. M. (2017). How embryophytic is the biosynthesis of phenylpropanoids and their derivatives in streptophyte algae? *Plant Cell Physiol.* 58, 934–945. doi: 10.1093/pcp/pcx037
- Delwiche, C. F., and Cooper, E. D. (2015). The evolutionary origin of a terrestrial flora. *Curr. Biol.* 25, R899–R910. doi: 10.1016/j.cub.2015.08.029
- Demarsy, E., Goldschmidt-Clermont, M., and Ulm, R. (2017). Coping with ‘Dark Sides of the Sun’ through photoreceptor signaling. *Trends Plant Sci.* 23, 260–271. doi: 10.1016/j.tplants.2017.11.007
- Dima, O., Morreel, K., Vanholme, B., Kim, H., Ralph, J., and Boerjan, W. (2015). Small glycosylated lignin oligomers are stored in Arabidopsis leaf vacuoles. *Plant Cell* 27, 695–710. doi: 10.1105/tpc.114.134643
- El Shoubaky, G. A., Abdel-Daim, M. M., Mansour, M. H., and Salem, E. A. (2016). Isolation and identification of a flavone apigenin from marine red alga *Acanthophora spicifera* with antinociceptive and anti-inflammatory activities. *J. Exp. Neurosci.* 10, 21–29. doi: 10.4137/JEN.S25096
- Elumalai, P., and Liu, H. L. (2011). Homology modeling and dynamics study of aureusidin synthase - an important enzyme in aurone biosynthesis of snapdragon flower. *Int. J. Biol. Macromol.* 49, 134–142. doi: 10.1016/j.ijbiomac.2011.03.017
- Emiliani, G., Fondi, M., Fani, R., and Gribaldo, S. (2009). A horizontal gene transfer at the origin of phenylpropanoid metabolism: a key adaptation of plants to land. *Biol. Direct* 4, 7. doi: 10.1186/1745-6150-4-7
- Farag, M. A., Deavours, B. E., De Fatima, A., Naoumkina, M., Dixon, R. A., and Sumner, L. W. (2009). Integrated metabolite and transcript profiling identify a biosynthetic mechanism for hispidol in *Medicago truncatula* cell cultures. *Plant Physiol.* 151, 1096–1113. doi: 10.1104/pp.109.141481
- Feller, A., Machemer, K., Braun, E. L., and Grotewold, E. (2011). Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *Plant J.* 66, 94–116. doi: 10.1111/j.1365-3113X.2010.04459.x
- Flores-Sandoval, E., Eklund, D. M., and Bowman, J. L. (2015). A simple auxin transcriptional response system regulates multiple morphogenetic processes in the liverwort *Marchantia polymorpha*. *PLoS Genet.* 11, e1005207. doi: 10.1371/journal.pgen.1005207
- Foyer, C. H. (2018). Reactive oxygen species, oxidative signaling and the regulation of photosynthesis. *Environ. Exp. Bot.* 154, 134–142. doi: 10.1016/j.envexpbot.2018.05.003
- Gang, D. R., Costa, M. A., Fujita, M., Dinkova-Kostova, A. T., Wang, H. B., Burlat, V., et al. (1999). Regiochemical control of monolignol radical coupling: a new paradigm for lignin and lignan biosynthesis. *Chem. Biol.* 6, 143–151. doi: 10.1016/S1074-5521(99)89006-1
- Gasparrini, C., Faccio, G., Arvas, M., Buchert, J., Saloheimo, M., and Kruus, K. (2010). Discovery of a new tyrosinase-like enzyme family lacking a C-terminally processed domain: production and characterization of an *Aspergillus oryzae* catechol oxidase. *Appl. Microbiol. Biotechnol.* 86, 213–226. doi: 10.1007/s00253-009-2258-3
- Gasper, R., Effenberger, I., Kolesinski, P., Terlecka, B., Hofmann, E., and Schaller, A. (2016). Dirigent protein mode of action revealed by the crystal structure of AtDIR6. *Plant Physiol.* 172, 2165–2175. doi: 10.1104/pp.16.01281
- Gerdol, R., Bonora, A., Marchesini, R., Gualandri, R., and Pancaldi, S. (1998). Growth response of *Sphagnum capillifolium* to nighttime temperature and nutrient level: Mechanisms and implications for global change. *Arct. Alpine Res.* 30, 388–395. doi: 10.2307/1552011
- Gerdol, R. (1996). The seasonal growth pattern of *Sphagnum magellanicum* Brid. in different microhabitats on a mire in the southern Alps (Italy). *Oecologia Montana* 5, 13–20.
- Glime, J. M. (2007). *Bryophyte Ecology. Volume 1. Physiological Ecology* (Michigan Technological University and the International Association of Bryologists).
- Goiris, K., Muylaert, K., Voorspoels, S., Noten, B., Baart, G. J. E., and De Cooman, L. (2014). Detection of flavonoids in microalgae from different evolutionary lineages. *J. Phycol.* 50, 483–492. doi: 10.1111/jpy.12180
- Gould, K. S., Kuhn, D. N., Lee, D. W., and Oberbauer, S. F. (1995). Why leaves are sometimes red. *Nature* 378, 241–242. doi: 10.1038/378241b0
- Han, X. J., Wu, Y. F., Gao, S., Yu, H. N., Xu, R. X., Lou, H. X., et al. (2014). Functional characterization of a *Plagiochasma appendiculatum* flavone synthase I showing flavanone 2-hydroxylase activity. *FEBS Lett.* 588, 2307–2314. doi: 10.1016/j.febslet.2014.05.023
- Harholt, J., Moestrup, O., and Ulvskov, P. (2016). Why plants were terrestrial from the beginning. *Trends Plant Sci.* 21, 96–101. doi: 10.1016/j.tplants.2015.11.010
- Holzinger, A., and Pichrtová, M. (2016). Abiotic stress tolerance of Charophyte green algae: new challenges for Omics techniques. *Front. Plant Sci.* 7, 678. doi: 10.3389/fpls.2016.00678
- Hong, L., Qian, Q., Tang, D., Wang, K., Li, M., and Cheng, Z. (2012). A mutation in the rice chalcone isomerase gene causes the *golden hull* and *internode 1* phenotype. *Planta* 236, 141–151. doi: 10.1007/s00425-012-1598-x
- Hooijmaijers, C. A. M., and Gould, K. S. (2007). Photoprotective pigments in red and green gametophytes of two New Zealand liverworts. *N. Z. J. Bot.* 45, 451–461. doi: 10.1080/00288250709509728
- Hosmani, P. S., Kamiya, T., Danku, J., Naseer, S., Geldner, N., Guerinot, M. L., et al. (2013). Dirigent domain-containing protein is part of the machinery required for formation of the lignin-based Casparian strip in the root. *Proc. Natl. Acad. Sci. U.S.A.* 110, 14498–14503. doi: 10.1073/pnas.1308412110
- Huber, M., Hintermann, G., and Lerch, K. (1985). Primary structure of tyrosinase from *Streptomyces glaucescens*. *Biochemistry* 24, 6038–6044. doi: 10.1021/bi00343a003
- Hughes, N. M. (2011). Winter leaf reddening in ‘evergreen’ species. *New Phytol.* 190, 573–581. doi: 10.1111/j.1469-8137.2011.03662.x
- Hussain, T., Plunkett, B., Ejaz, M., Espley, R. V., and Kayser, O. (2018). Identification of putative precursor genes for the biosynthesis of cannabinoid-like compound in *Radula marginata*. *Front. Plant Sci.* 9, 537. doi: 10.3389/fpls.2018.00537
- Imbs, T. I., and Zvyagintseva, T. N. (2018). Phlorotannins are polyphenolic metabolites of brown algae. *Russ. J. Mar. Biol.* 44, 263–273. doi: 10.1134/S106307401804003X
- Ishizaki, K., Nishihama, R., Yamato, K. T., and Kohchi, T. (2015). Molecular genetic tools and techniques for *Marchantia polymorpha* research. *Plant Cell Physiol.* 57, 262–270. doi: 10.1093/pcp/pcv097
- Jez, J. M., and Noel, J. P. (2002). Reaction Mechanism of chalcone isomerase: pH dependence, diffusion control, and product binding differences. *J. Biol. Chem.* 277, 1361–1369. doi: 10.1074/jbc.M109224200

- Jiang, W., Yin, Q., Wu, R., Zheng, G., Liu, J., Dixon, R. A., et al. (2015). Role of a chalcone isomerase-like protein in flavonoid biosynthesis in *Arabidopsis thaliana*. *J. Exp. Bot.* 66, 7165–7179. doi: 10.1093/jxb/erv413
- Jiang, N., Doseff, A. I., and Grotewold, E. (2016). Flavones: From biosynthesis to health benefits. *Plants* 5, E27. doi: 10.3390/plants502027
- Jin, H., Cominelli, E., Bailey, P., Parr, A., Mehrtens, F., Jones, J., et al. (2000). Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in *Arabidopsis*. *EMBO J.* 19, 6150–6161. doi: 10.1093/emboj/19.22.6150
- Jorgensen, R. (1994). “The genetic origins of biosynthesis and light-responsive control of the chemical UV screen of land plants,” in *Recent Advances in Phytochemistry*, vol. 28. Eds. B. E. Ellis, G. W. Kuroki and H. A. Stafford (Boston, USA: Springer), 179–192. doi: 10.1007/978-1-4615-2544-8_7
- Kaintz, C., Molitor, C., Thill, J., Kampatsikas, I., Michael, C., Halbwirth, H., et al. (2014). Cloning and functional expression in *E. coli* of a polyphenol oxidase transcript from *Coreopsis grandiflora* involved in aurone formation. *FEBS Lett.* 588, 3417–3426. doi: 10.1016/j.febslet.2014.07.034
- Kaltenbach, M., Burke, J. R., Dindo, M., Pabis, A., Munsberg, F. S., Rabin, A., et al. (2018). Evolution of chalcone isomerase from a noncatalytic ancestor. *Nat. Chem. Biol.* 14, 988. doi: 10.1038/s41589-018-0079-310.1038/s41589-018-0042-3
- Karsten, U., and Holzinger, A. (2014). Green algae in alpine biological soil crust communities: acclimation strategies against ultraviolet radiation and dehydration. *Biodivers. Conserv.* 23, 1845–1858. doi: 10.1007/s10531-014-0653-2
- Kenrick, P., and Crane, P. R. (1997). The origin and early evolution of plants on land. *Nature* 389, 33–39. doi: 10.1038/37918
- Kim, K. W., Smith, C. A., Daily, M. D., Cort, J. R., Davin, L. B., and Lewis, N. G. (2015). Trimeric structure of (+)-pinorensin-forming dirigent protein at 1.95 Å resolution with three isolated active sites. *J. Biol. Chem.* 290, 1308–1318. doi: 10.1074/jbc.M114.611780
- Klejdus, B., Lojkova, L., Plaza, M., Snoblova, M., and Sterbova, D. (2010). Hyphenated technique for the extraction and determination of isoflavones in algae: ultrasound-assisted supercritical fluid extraction followed by fast chromatography with tandem mass spectrometry. *J. Chromatogr. A* 1217, 7956–7965. doi: 10.1016/j.chroma.2010.07.020
- Kny, L. (1890). “Bau und Entwicklung von *Marchantia polymorpha*,” in *Botanische Wandtafeln, Vol VIII Abtheilung* (Berlin, Germany: Paul Parey), 364–401.
- Koes, R. E., Quattrocchio, F., and Mol, J. N. M. (1994). The flavonoid biosynthetic pathway in plants: function and evolution. *Bioessays* 16, 123–132. doi: 10.1002/bies.950160209
- Kondou, Y., Miyagi, Y., Morito, T., Fujihira, K., Miyauchi, W., Moriyama, A., et al. (2019). Physiological function of photoreceptor UVR8 in UV-B tolerance in the liverwort *Marchantia polymorpha*. *Planta* 249, 1349–1364. doi: 10.1007/s00425-019-03090-w
- Kubo, H., Nozawa, S., Hiwatashi, T., Kondou, Y., Nakabayashi, R., Mori, T., et al. (2018). Biosynthesis of riccionidins and marchantins is regulated by R2R3-MYB transcription factors in *Marchantia polymorpha*. *J. Plant Res.* 131, 849–864. doi: 10.1007/s10265-018-1044-7
- Kunz, S., and Becker, H. (1995). Cell wall pigment formation of *in vitro* cultures of the liverwort *Ricciocarpus natans*. *Z. für Naturforschung C* 50, 235–240. doi: 10.1515/znc-1995-3-412
- Kunz, S., Burkhardt, G., and Becker, H. (1993). Riccionidins A and B, anthocyanidins from the cell walls of the liverwort *Ricciocarpus natans*. *Phytochemistry* 35, 233–235. doi: 10.1016/S0031-9422(00)90540-5
- Kusano, M., Tohge, T., Fukushima, A., Kobayashi, M., Hayashi, N., Otsuki, H., et al. (2011). Metabolomics reveals comprehensive reprogramming involving two independent metabolic responses of *Arabidopsis* to UV-B light. *Plant J.* 67, 354–369. doi: 10.1111/j.1365-3113.2011.04599.x
- Labandeira, C., and Currano, E. (2013). The fossil record of plant-insect dynamics. *Ann. Rev. Earth Planet. Sci.* 41, 287–311. doi: 10.1146/annurev-earth-050212-124139
- Labeuw, L., Martone, P. T., Boucher, Y., and Case, R. J. (2015). Ancient origin of the biosynthesis of lignin precursors. *Biol. Direct* 10, 23. doi: 10.1186/s13062-015-0052-y
- Landi, M., Tattini, M., and Gould, K. S. (2015). Multiple functional roles of anthocyanins in plant-environment interactions. *Environ. Exp. Bot.* 119, 4–17. doi: 10.1016/j.envexpbot.2015.05.012
- Landry, L. G., Chapple, C. C., and Last, R. L. (1995). *Arabidopsis* mutants lacking phenolic sunscreens exhibit enhanced ultraviolet-B injury and oxidative damage. *Plant Physiol.* 109, 1159–1166. doi: 10.1104/pp.109.4.1159
- Lee, D. W., and Gould, K. S. (2002). Anthocyanins in leaves and other vegetative organs: An introduction. *Adv. Bot. Res.* 37, 1–16. doi: 10.1016/S0065-2296(02)37040-X
- Li, J., Ou-Lee, M., Amundson, R. G., and Last, R. L. (1993). *Arabidopsis* flavonoid mutants are hypersensitive to UV-B irradiation. *Plant Cell* 5, 171–179. doi: 10.2307/386958310.1105/tpc.5.2.171
- Li, C., Liu, S., Zhang, W., Chen, K., and Zhang, P. (2019). Transcriptional profiling and physiological analysis reveal the critical roles of ROS-scavenging system in the Antarctic moss *Pohlia nutans* under Ultraviolet-B radiation. *Plant Physiol. Biochem.* 134, 113–122. doi: 10.1016/j.plaphy.2018.10.034
- Ligrone, R., Duckett, J. G., and Renzaglia, K. S. (2012). Major transitions in the evolution of early land plants: a bryological perspective. *Ann. Bot.* 109, 851–871. doi: 10.1093/aob/mcs017
- Lillo, C., Lea, U. S., and Ruoff, P. (2008). Nutrient depletion as a key factor for manipulating gene expression and product formation in different branches of the flavonoid pathway. *Plant Cell Environ.* 31, 587–601. doi: 10.1111/j.1365-3040.2007.01748.x
- Lindenberg, J. B. W. (1836). *Monographie der Riccieen. Nova Acta Physico-Medica Academiae Caesareae Leopoldino - Carolinae Naturae Curiosum XVIII*, 361–504.
- Lloyd, A., Brockman, A., Aguirre, L., Campbell, A., Bean, A., Cantero, A., et al. (2017). Advances in the MYB-bHLH-WD Repeat (MBW) pigment regulatory model: Addition of a WRKY factor and co-option of an anthocyanin MYB for betalain regulation. *Plant Cell Physiol.* 58, 1431–1441. doi: 10.1093/pcp/pcx075
- Lu, Y., Ye, C., Che, J., Xu, X., Shao, D., Jiang, C., et al. (2019). Genomic sequencing, genome-scale metabolic network reconstruction, and in silico flux analysis of the grape endophytic fungus *Alternaria* sp. MG1. *Microbial Cell Fact.* 18, 13. doi: 10.1186/s12934-019-1063-7
- Ma, D., and Constabel, C. P. (2019). MYB repressors as regulators of phenylpropanoid metabolism in plants. *Trends Plant Sci.* 24, 275–289. doi: 10.1016/j.tplants.2018.12.003
- Maloney, G. S., DiNapoli, K. T., and Muday, G. K. (2014). The *anthocyanin reduced* tomato mutant demonstrates the role of flavonols in tomato lateral root and root hair development. *Plant Physiol.* 166, 614–631. doi: 10.1104/pp.114.240507
- Manetas, Y. (2006). Why some leaves are anthocyanic and why most anthocyanic leaves are red? *Flora* 201, 163–177. doi: 10.1016/j.flora.2005.06.010
- Markham, K. R. (1988). “Distribution of flavonoids in the lower plants and its evolutionary significance,” in *The Flavonoids*. Ed. J. B. Harborne (London, UK: Chapman and Hall), 427–468. doi: 10.1007/978-1-4899-2913-6_12
- Martens, S., Preuß, A., and Matern, U. (2010). Multifunctional flavonoid dioxygenases: Flavonol and anthocyanin biosynthesis in *Arabidopsis thaliana* L. *Phytochemistry* 71, 1040–1049. doi: 10.1016/j.phytochem.2010.04.016
- Matoba, Y., Kumagai, T., Yamamoto, A., Yoshitsu, H., and Sugiyama, M. (2006). Crystallographic evidence that the dinuclear copper center of tyrosinase is flexible during catalysis. *J. Biol. Chem.* 281, 8981–8990. doi: 10.1074/jbc.M509785200
- Merzlyak, M. N., and Chivkunova, O. B. (2000). Light-stress-induced pigment changes and evidence for anthocyanin photoprotection in apples. *J. Photochem. Photobiol. B Biol.* 55, 155–163. doi: 10.1016/S1011-1344(00)00042-7
- Meslet-Cladière, L., Delage, L., Leroux, C. J.-J., Goullitquer, S., Leblanc, C., Creis, E., et al. (2013). Structure/function analysis of a Type III polyketide synthase in the brown alga *Ectocarpus siliculosus* reveals a biochemical pathway in phlorotannin monomer biosynthesis. *Plant Cell* 25, 3089–3103. doi: 10.1105/tpc.113.111336
- Miyahara, T., Sugishita, N., Ishida-Dei, M., Okamoto, E., Kouno, T., Cano, E. A., et al. (2018). Carnation *I* locus contains two chalcone isomerase genes involved in orange flower coloration. *Breed. Sci.* 68, 481–487. doi: 10.1270/jsbbs.18029
- Molitor, C., Mauracher, S. G., Pargan, S., Mayer, R. L., Halbwirth, H., and Rempel, A. (2015). Latent and active aurone synthase from petals of *C. grandiflora*: a

- polyphenol oxidase with unique characteristics. *Planta* 242, 519–537. doi: 10.1007/s00425-015-2261-0
- Molitor, C., Mauracher, S. G., and Rempel, A. (2016). Aurone synthase is a catechol oxidase with hydroxylase activity and provides insights into the mechanism of plant polyphenol oxidases. *Proc. Nat. Acad. Sci. U.S.A.* 113, E1806–E1815. doi: 10.1073/pnas.1523575113
- Monforte, L., Soriano, G., Núñez-Olivera, E., and Martínez-Abaigar, J. (2018). Cell compartmentation of ultraviolet-absorbing compounds: An underexplored tool related to bryophyte ecology, phylogeny and evolution. *Funct. Ecol.* 32, 882–893. doi: 10.1111/1365-2435.13048
- Morales, L. O., Brosché, M., Vainonen, J., Jenkins, G. I., Wargent, J. J., Sipari, N., et al. (2013). Multiple roles for UV RESISTANCE LOCUS8 in regulating gene expression and metabolite accumulation in *Arabidopsis* under solar ultraviolet radiation. *Plant Physiol.* 161, 744–759. doi: 10.1104/pp.112.211375
- Mouradov, A., and Spangenberg, G. (2014). Flavonoids: a metabolic network mediating plants adaptation to their real estate. *Front. Plant Sci.* 5, 620. doi: 10.3389/fpls.2014.00620
- Muhlemann, J. K., Younts, T. L. B., and Muday, G. K. (2018). Flavonols control pollen tube growth and integrity by regulating ROS homeostasis during high-temperature stress. *Proc. Nat. Acad. Sci. U.S.A.* 115, E11188–E11199. doi: 10.1073/pnas.1811492115
- Myburg, A. A., Grattapaglia, D., Tuskan, G. A., Hellsten, U., Hayes, R. D., Grimwood, J., et al. (2014). The genome of *Eucalyptus grandis*. *Nature* 510, 356–362. doi: 10.1038/nature13308
- Nagai, I. (1915). Ueber roten pigmentbildung bei einigen *Marchantia*-Arten. *Bot. Mag.* 29, 90–98. doi: 10.15281/jplantres1887.29.342_19910.15281/jplantres1887.29.342_90
- Nakayama, T., Yonekura-Sakakibara, K., Sato, T., Kikuchi, S., Fukui, Y., Fukuchi-Mizutani, M., et al. (2000). Aureusidin synthase: a polyphenol oxidase homolog responsible for flower coloration. *Science* 290, 1163–1166. doi: 10.1126/science.290.5494.1163
- Napoli, C. A., Fahy, D., Wang, H. Y., and Taylor, L. P. (1999). *white anther*: a *Petunia* mutant that abolishes pollen flavonol accumulation, induces male sterility, and is complemented by a chalcone synthase transgene. *Plant Physiol.* 120, 615–622. doi: 10.1104/pp.120.2.615
- Nemesio-Gorritz, M., Blair, P. B., Dalman, K., Hammerbacher, A., Arnerup, J., Stenlid, J., et al. (2017). Identification of Norway Spruce MYB-bHLH-WDR transcription factor complex members linked to regulation of the flavonoid pathway. *Front. Plant Sci.* 8, 305–305. doi: 10.3389/fpls.2017.00305
- Newsham, K. K., Geissler, P. A., Nicolson, M. J., Peat, H. J., and Lewis-Smith, R. I. (2005). Sequential reduction of UV-B radiation in the field alters the pigmentation of an Antarctic leafy liverwort. *Environ. Exp. Bot.* 54, 22–32. doi: 10.1016/j.envexpbot.2004.05.006
- Ngaki, M. N., Louie, G. V., Philippe, R. N., Manning, G., Pojer, F., Bowman, M. E., et al. (2012). Evolution of the chalcone-isomerase fold from fatty-acid binding to stereospecific catalysis. *Nature* 485, 530–533. doi: 10.1038/nature11009
- Ono, E., Hatayama, M., Isono, Y., Sato, T., Watanabe, R., Yonekura-Sakakibara, K., et al. (2006). Localization of a flavonoid biosynthetic polyphenol oxidase in vacuoles. *Plant J.* 45, 133–143. doi: 10.1111/j.1365-3113X.2005.02625.x
- Pandith, S. A., Ramazan, S., Khan, M. I., Reshi, Z. A., and Shah, M. A. (2020). Chalcone synthases (CHSs): the symbolic type III polyketide synthases. *Planta* 251, 15. doi: 10.1007/s00425-019-03307-y
- Paniagua, C., Bilkova, A., Jackson, P., Dabrowski, S., Riber, W., Didi, V., et al. (2017). Dirigent proteins in plants: modulating cell wall metabolism during abiotic and biotic stress exposure. *J. Exp. Bot.* 68, 3287–3301. doi: 10.1093/jxb/erx141
- Peer, W. A., and Murphy, A. S. (2007). Flavonoids and auxin transport: modulators or regulators? *Trends Plant Sci.* 12, 556–563. doi: 10.1016/j.tplants.2007.10.003
- Petersen, M., and Simmonds, M. S. (2003). Rosmarinic acid. *Phytochemistry* 62, 121–125. doi: 10.1016/S0031-9422(02)00513-7
- Philpott, M., Ferguson, L. R., Gould, K. S., and Harris, P. J. (2009). Anthocyanidin-containing compounds occur in the periderm cell walls of the storage roots of sweet potato (*Ipomoea batatas*). *J. Plant Physiol.* 166, 1112–1117. doi: 10.1016/j.jplph.2008.12.007
- Pichersky, E., and Gang, D. R. (2000). Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. *Trends Plant Sci.* 5, 439–445. doi: 10.1016/S1360-1385(00)01741-6
- Pietrini, F., Iannelli, M. A., and Massacci, A. (2002). Anthocyanin accumulation in the illuminated surface of maize leaves enhances protection from photo-inhibitory risks at low temperature, without further limitation to photosynthesis. *Plant Cell Environ.* 25, 1251–1259. doi: 10.1046/j.1365-3040.2002.00917.x
- Pollastri, S., and Tattini, M. (2011). Flavonols: old compounds for old roles. *Ann. Bot.* 108, 1225–1233. doi: 10.1093/aob/mcr234
- Polturak, G., and Aharoni, A. (2019). Advances and future directions in betalain metabolic engineering. *New Phytol.* 224, 1472–1478. doi: 10.1111/nph.15973
- Puttick, M. N., Morris, J. L., Williams, T. A., Cox, C. J., Edwards, D., Kenrick, P., et al. (2018). The interrelationships of land plants and the nature of the ancestral embryophyte. *Curr. Biol.* 28, 733–745. doi: 10.1016/j.cub.2018.01.063
- Ramos, P., Guajardo, J., Moya-León, M. A., and Herrera, R. (2016). A differential distribution of auxin and flavonols in radiata pine stem seedlings exposed to inclination. *Tree Genet. Genomes* 12, 42. doi: 10.1007/s11295-016-1003-1
- Reboledo, G., Del Campo, R., Alvarez, A., Montesano, M., Mara, H., and Ponce De Leon, I. (2015). *Physcomitrella patens* activates defense responses against the pathogen *Colletotrichum gloeosporioides*. *Int. J. Mol. Sci.* 16, 22280–22298. doi: 10.3390/ijms160922280
- Reeb, C., Marline, L., Rabeau, L., Andriamanantena, A., Andriamiarisoa, R. L., Ranarijaona, H., et al. (2018). A survey of Marchantiales from Madagascar. *Acta Biol. Plant Agriensis* 6, 3–72. doi: 10.21406/abpa.2018.6.3
- Rensing, S. A. (2018). Great moments in evolution: the conquest of land by plants. *Curr. Opin. Plant Biol.* 42, 49–54. doi: 10.1016/j.pbi.2018.02.006
- Rozema, J., Bjorn, L. O., Bornman, J. F., Gaberscik, A., Hader, D. P., Trost, T., et al. (2002). The role of UV-B radiation in aquatic and terrestrial ecosystems—an experimental and functional analysis of the evolution of UV-absorbing compounds. *J. Photochem. Photobiol. B: Biol.* 66, 2–12. doi: 10.1016/S1011-1344(01)00269-X
- Rubin, G., Tohge, T., Matsuda, F., Saito, K., and Scheible, W. R. (2009). Members of the LBD family of transcription factors repress anthocyanin synthesis and affect additional nitrogen responses in *Arabidopsis*. *Plant Cell* 21, 3567–3584. doi: 10.1105/tpc.109.067041
- Rudolph, H., Kabsch, U., and Schmidt-Stohn, G. (1977). Änderungen des Chloroplastenpigment-Spiegels bei *Sphagnum magellanicum* im Verlauf der Synthese von Sphagnorubin und anderer membranochromer Pigmente. *Z. für Pflanzenphysiologie* 82, 107–116. doi: 10.1016/S0044-328X(77)80219-5
- Sendovski, M., Kantev, M., Ben-Yosef, V. S., Adir, N., and Fishman, A. (2011). First structures of an active bacterial tyrosinase reveal copper plasticity. *J. Mol. Biol.* 405, 227–237. doi: 10.1016/j.jmb.2010.10.048
- Shimamura, M. (2016). *Marchantia polymorpha*: Taxonomy, phylogeny and morphology of a model system. *Plant Cell Physiol.* 57, 230–256. doi: 10.1093/pcp/pcv192
- Shuster, V., and Fishman, A. (2009). Isolation, cloning and characterization of a tyrosinase with improved activity in organic solvents from *Bacillus megaterium*. *J. Mol. Microbiol. Biotechnol.* 17, 188–200. doi: 10.1159/000233506
- Soriano, G., Cloix, C., Heilmann, M., Nunez-Olivera, E., Martínez-Abaigar, J., and Jenkins, G. I. (2018a). Evolutionary conservation of structure and function of the UVR8 photoreceptor from the liverwort *Marchantia polymorpha* and the moss *Physcomitrella patens*. *New Phytol.* 217, 151–162. doi: 10.1111/nph.14767
- Soriano, G., Del-Castillo-Alonso, M. A., Monforte, L., Nunez-Olivera, E., and Martínez-Abaigar, J. (2018b). Phenolic compounds from different bryophyte species and cell compartments respond specifically to ultraviolet radiation, but not particularly quickly. *Plant Physiol. Biochem.* 134, 137–144. doi: 10.1016/j.plaphy.2018.07.020
- Soriano, G., Del-Castillo-Alonso, M. A., Monforte, L., Nunez-Olivera, E., and Martínez-Abaigar, J. (2018c). First data on the effects of ultraviolet radiation on phenolic compounds in the model hornwort *Anthoceros agrestis*. *Cryptogamie Bryologie* 39, 201–211. doi: 10.7872/cryb/v39.iss2.2018.201
- Stafford, H. A. (1991). Flavonoid evolution: An enzymic approach. *Plant Physiol.* 96, 680–685. doi: 10.1104/pp.96.3.680
- Stark, L. R., Greenwood, J. L., and Brinda, J. C. (2017). Desiccated *Syntrichia ruralis* shoots regenerate after 20 years in the herbarium. *J. Bryol.* 39, 85–93. doi: 10.1080/03736687.2016.1176307
- Stich, K., Eidenberger, T., Wurst, F., and Forkmann, G. (1992). Enzymatic conversion of dihydroflavonols to flavan-3,4-diols using flower extracts of

- Dianthus caryophyllus* L. (carnation). *Planta* 187, 103–108. doi: 10.1007/BF00201630
- Sugano, S. S., Nishihama, R., Shirakawa, M., Takagi, J., Matsuda, Y., Ishida, S., et al. (2018). Efficient CRISPR/Cas9-based genome editing and its application to conditional genetic analysis in *Marchantia polymorpha*. *PLoS One* 13, e0205117. doi: 10.1371/journal.pone.0205117
- Szövényi, P., Frangedakis, E., Ricca, M., Quandt, D., Wicke, S., and Langdale, J. A. (2015). Establishment of *Anthoceros agrestis* as a model species for studying the biology of hornworts. *BMC Plant Biol.* 15, 98. doi: 10.1186/s12870-015-0481-x
- Taniguchi, S., Yazaki, K., Yabuuchi, R., Kawakami, K., Ito, H., Hatano, T., et al. (2000). Galloylglucoses and riccionidin A in *Rhus javanica* adventitious root cultures. *Phytochemistry* 53, 357–363. doi: 10.1016/S0031-9422(99)00399-4
- Taylor, L. P., and Grotewold, E. (2005). Flavonoids as developmental regulators. *Curr. Opin. Plant Biol.* 8, 317–323. doi: 10.1016/j.pbi.2005.03.005
- Thamil Arasan, S. K., Park, J. I., Ahmed, N. U., Jung, H. J., Hur, Y., Kang, K. K., et al. (2013). Characterization and expression analysis of dirigent family genes related to stresses in Brassica. *Plant Physiol. Biochem.* 67, 144–153. doi: 10.1016/j.plaphy.2013.02.030
- Tohge, T., Watanabe, M., Hoefgen, R., and Fernie, A. R. (2013). The evolution of phenylpropanoid metabolism in the green lineage. *Crit. Rev. Biochem. Mol. Biol.* 48, 123–152. doi: 10.3109/10409238.2012.758083
- Tran, L. T., Taylor, J. S., and Constabel, C. P. (2012). The polyphenol oxidase gene family in land plants: Lineage-specific duplication and expansion. *BMC Genomics* 13, 395. doi: 10.1186/1471-2164-13-395
- Uchida, K., Akashi, T., and Aoki, T. (2017). The missing link in Leguminous pterocarpan biosynthesis is a dirigent domain-containing protein with isoflavanol dehydratase activity. *Plant Cell Physiol.* 58, 398–408. doi: 10.1093/pcp/pcw213
- Van der Meer, I. M., Stam, M. E., Van Tunen, A. J., Mol, J. N., and Stuitje, A. R. (1992). Antisense inhibition of flavonoid biosynthesis in *Petunia* anthers results in male sterility. *Plant Cell* 4, 253–262. doi: 10.1105/tpc.4.3.253
- Verhoeven, J. T. A., and Liefveld, W. M. (1997). The ecological significance of organochemical compounds in *Sphagnum*. *Acta Botanica Neerlandica* 46, 117–130. doi: 10.1111/plb.1997.46.2.117
- Vitt, D. H., Crandall-Stotler, B. J., and Wood, A. (2012). “Bryophytes: survival in a dry world through tolerance and avoidance,” in *Plant Ecology and Evolution in Harsh Environments*. Eds. N. Rajakaruna, R. S. Boyd and T. B. Harris (New York, USA: Nova Science), 267–295.
- Vowinkel, E. (1975). Torfmoosmembranochrome, 2. Die struktur des sphagnorubins. *Chemische Berichte-Recueil* 108, 1166–1181. doi: 10.1002/cber.19751080423
- Wang, Y., Chantreau, M., Sibout, R., and Hawkins, S. (2013). Plant cell wall lignification and monolignol metabolism. *Front. Plant Sci.* 4, 220. doi: 10.3389/fpls.2013.00220
- Wang, N., Liu, W., Zhang, T., Jiang, S., Xu, H., Wang, Y., et al. (2018). Transcriptomic analysis of red-fleshed apples reveals the novel role of MdWRKY11 in flavonoid and anthocyanin biosynthesis. *J. Agric. Food Chem.* 66, 7076–7086. doi: 10.1021/acs.jafc.8b01273
- Waterman, M. J., Nugraha, A. S., Hendra, R., Ball, G. E., Robinson, S. A., and Keller, P. A. (2017). Antarctic moss biflavonoids show high antioxidant and ultraviolet-screening activity. *J. Nat. Prod.* 80, 2224–2231. doi: 10.1021/acs.jnatprod.7b00085
- Waterman, M. J., Bramley-Alves, J., Miller, R. E., Keller, P. A., and Robinson, S. A. (2018). Photoprotection enhanced by red cell wall pigments in three East Antarctic mosses. *Biol. Res.* 51, 49. doi: 10.1186/s40659-018-0196-1
- Wickett, N. J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., and Matasci, N. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc. Natl. Acad. Sci. U.S.A.* 111, E4859–E4868. doi: 10.1073/pnas.1323926111
- Wolf, L., Rizzini, L., Stracke, R., Ulm, R., and Rensing, S. A. (2010). The molecular and physiological responses of *Physcomitrella patens* to Ultraviolet-B radiation. *Plant Physiol.* 153, 1123–1134. doi: 10.1104/pp.110.154658
- Wu, Y. F., Zhao, Y., Liu, X. Y., Gao, S., Cheng, A. X., and Lou, H. X. (2018). A bHLH transcription factor regulates bisbenzyl biosynthesis in the liverwort *Plagiochasma appendiculatum*. *Plant Cell Physiol.* 59, 1187–1199. doi: 10.1093/pcp/pcy053
- Yin, R., and Ulm, R. (2017). How plants cope with UV-B: from perception to response. *Curr. Opin. Plant Biol.* 37, 42–48. doi: 10.1016/j.pbi.2017.03.013
- Ylstra, B., Muskens, M., and Van Tunen, A. J. (1996). Flavonols are not essential for fertilization in *Arabidopsis thaliana*. *Plant Mol. Biol.* 32, 1155–1158. doi: 10.1007/BF00041399
- Yonekura-Sakakibara, K., Higashi, Y., and Nakabayashi, R. (2019). The origin and evolution of plant flavonoid metabolism. *Front. Plant Sci.* 10, 943. doi: 10.3389/fpls.2019.00943
- Zhao, Q., and Dixon, R. A. (2014). Altering the cell wall and its impact on plant disease: from forage to bioenergy. *Ann. Rev. Phytopathol.* 52, 69–91. doi: 10.1146/annurev-phyto-082712-102237
- Zhao, D., Hamilton, J. P., Bhat, W. W., Johnson, S. R., Godden, G. T., Kinser, T. J., et al. (2019). A chromosomal-scale genome assembly of *Tectona grandis* reveals the importance of tandem gene duplication and enables discovery of genes in natural product biosynthetic pathways. *GigaScience* 8, giz005. doi: 10.1093/gigascience/giz005
- Zhong, B., Xi, Z., Goremykin, V. V., Fong, R., Mclenachan, P. A., Novis, P. M., et al. (2014). Streptophyte algae and the origin of land plants revisited using heterogeneous models with three new algal chloroplast genomes. *Mol. Biol. Evol.* 31, 177–183. doi: 10.1093/molbev/mst200
- Zimmermann, I. M., Heim, M. A., Weisshaar, B., and Uhrig, J. F. (2004). Comprehensive identification of *Arabidopsis thaliana* MYB transcription factors interacting with R/B-like BHLH proteins. *Plant J.* 40, 22–34. doi: 10.1111/j.1365-313X.2004.02183.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Davies, Jibran, Zhou, Albert, Brummell, Jordan, Bowman and Schwinn. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read
for greatest visibility
and readership



FAST PUBLICATION

Around 90 days
from submission
to decision



HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,
and constructive
peer-review



TRANSPARENT PEER-REVIEW

Editors and reviewers
acknowledged by name
on published articles

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org

Contact us: info@frontiersin.org | +41 21 510 17 00



REPRODUCIBILITY OF RESEARCH

Support open data
and methods to enhance
research reproducibility



DIGITAL PUBLISHING

Articles designed
for optimal readership
across devices



FOLLOW US

[@frontiersin](https://twitter.com/frontiersin)



IMPACT METRICS

Advanced article metrics
track visibility across
digital media



EXTENSIVE PROMOTION

Marketing
and promotion
of impactful research



LOOP RESEARCH NETWORK

Our network
increases your
article's readership