

GENETIC REGULATORY MECHANISMS UNDERLYING DEVELOPMENTAL SHIFTS IN PLANT EVOLUTION

EDITED BY: Verónica S. Di Stilio, Annette Becker and Natalia Pabón-Mora
PUBLISHED IN: *Frontiers in Plant Science*





frontiers

Frontiers Copyright Statement

© Copyright 2007-2019 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88963-059-2

DOI 10.3389/978-2-88963-059-2

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

GENETIC REGULATORY MECHANISMS UNDERLYING DEVELOPMENTAL SHIFTS IN PLANT EVOLUTION

Topic Editors:

Verónica S. Di Stilio, University of Washington, United States

Annette Becker, Justus-Liebig-Universität Gießen, Germany

Natalia Pabón-Mora, Universidad de Antioquia, Colombia



Image by C. Scutt, A. Sengupta, V. Di Stilio and N. Pabon-Mora

Citation: Di Stilio, V. S., Becker, A., Pabón-Mora, N., eds. (2019). Genetic Regulatory Mechanisms Underlying Developmental Shifts in Plant Evolution. Lausanne: Frontiers Media. doi: 10.3389/978-2-88963-059-2

Table of Contents

05 Editorial: Genetic Regulatory Mechanisms Underlying Developmental Shifts in Plant Evolution

Natalia Pabón-Mora, Verónica S. Di Stilio and Annette Becker

ORIGINS OF COMPLEXITY: THEORETICAL APPROACHES AND THE STUDY OF GENE EVOLUTION OVER DEEP PHYLOGENETIC TIMESCALES

09 Dynamical Patterning Modules, Biogeneric Materials, and the Evolution of Multicellular Plants

Mariana Benítez, Valeria Hernández-Hernández, Stuart A. Newman and Karl J. Niklas

25 Evolution of Protein Ductility in Duplicated Genes of Plants

Inmaculada Yruela, Bruno Contreras-Moreira, A. Keith Dunker and Karl J. Niklas

35 DNA Methylation and the Evolution of Developmental Complexity in Plants

Katharina Bräutigam and Quentin Cronk

49 The Times They are A-Changin': Heterochrony in Plant Development and Evolution

Manuel Buendía-Monreal and C. Stewart Gillmor

57 A Phylogenetic Study of the ANT Family Points to a preANT Gene as the Ancestor of Basal and euANT Transcription Factors in Land Plants

Melissa Dipp-Álvarez and Alfredo Cruz-Ramírez

ORIGIN AND PATTERNING OF THE GAMETOPHYTE AND SPOROPHYTE IN EARLY-DIVERGING LAND PLANTS

69 Co-expression and Transcriptome Analysis of *Marchantia polymorpha* Transcription Factors Supports Class C ARFs as Independent Actors of an Ancient Auxin Regulatory Module

Eduardo Flores-Sandoval, Facundo Romani and John L. Bowman

90 Evolution of the Symbiosis-Specific GRAS Regulatory Network in Bryophytes

Christopher Grosche, Anne Christina Genau and Stefan A. Rensing

103 Getting to the Roots: A Developmental Genetic View of Root Anatomy and Function From *Arabidopsis* to *Lycophytes*

Frauke Augstein and Annelie Carlsbecker

EVOLUTION OF THE MEGASPORANGIUM (OVULE) AND MEGAGAMETOPHYTE IN SEED PLANTS

118 Activation of Nucleases, PCD, and Mobilization of Reserves in the *Araucaria angustifolia* Megagametophyte During Germination

Laura Moyano, María D. Correa, Leonardo C. Favre, Florencia S. Rodríguez, Sara Maldonado and María P. López-Fernández

- 131** *Evidence for the Extensive Conservation of Mechanisms of Ovule Integument Development Since the Most Recent Common Ancestor of Living Angiosperms*
 Gontran Arnault, Aurélie C. M. Vialette, Amélie Andres-Robin, Bruno Fogliani, Gildas Gâteblé and Charles P. Scutt
- 140** *Integrative Analysis of Three RNA Sequencing Methods Identifies Mutually Exclusive Exons of MADS-Box Isoforms During Early Bud Development in Picea abies*
 Shirin Akhter, Warren W. Kretzschmar, Veronika Nordal, Nicolas Delhomme, Nathaniel R. Street, Ove Nilsson, Olof Emanuelsson and Jens F. Sundström

EVOLUTION OF MOLECULAR MECHANISMS UNDERLYING FLOWERING

- 158** *The Functional Change and Deletion of FLC Homologs Contribute to the Evolution of Rapid Flowering in Boechera stricta*
 Cheng-Ruei Lee, Jo-Wei Hsieh, M. E. Schranz and Thomas Mitchell-Olds
- 169** *How to Evolve a Perianth: A Review of Cadastral Mechanisms for Perianth Identity*
 Marie Monniaux and Michiel Vandenbussche
- 176** *Gene Duplication and Transference of Function in the paleoAP3 Lineage of Floral Organ Identity Genes*
 Kelsey D. Galimba, Jesús Martínez-Gómez and Verónica S. Di Stilio
- 191** *Unraveling the Developmental and Genetic Mechanisms Underpinning Floral Architecture in Proteaceae*
 Catherine Damerval, Hélène Citerne, Natalia Conde e Silva, Yves Deveau, Etienne Delannoy, Johann Joets, Franck Simonnet, Yannick Staedler, Jürg Schönenberger, Jennifer Yansouni, Martine Le Guilloux, Hervé Sauquet and Sophie Nadot
- 206** *Novel Traits, Flower Symmetry, and Transcriptional Autoregulation: New Hypotheses From Bioinformatic and Experimental Data*
 Aniket Sengupta and Lena C. Hileman

EVOLUTION OF THE GENETIC NETWORK CONTROLLING FRUIT DEVELOPMENT

- 221** *Duplication and Diversification of REPLUMLESS – A Case Study in the Papaveraceae*
 Cecilia Zumajo-Cardona, Natalia Pabón-Mora and Barbara A. Ambrose
- 240** *Evolution and Diversification of FRUITFULL Genes in Solanaceae*
 Dinusha C. Maheepala, Christopher A. Emerling, Alex Rajewski, Jenna Macon, Maya Strahl, Natalia Pabón-Mora and Amy Litt



Editorial: Genetic Regulatory Mechanisms Underlying Developmental Shifts in Plant Evolution

Natalia Pabón-Mora^{1*}, Verónica S. Di Stilio² and Annette Becker³

¹ Instituto de Biología, Universidad de Antioquia, Medellín, Colombia, ² Department of Biology, University of Washington, Seattle, WA, United States, ³ Institute of Botany, Justus-Liebig-Universität Gießen, Giessen, Germany

Keywords: plant development, evo devo, plant evolution, reproductive strategies, developmental modularity

Editorial on the Research Topic

Genetic Regulatory Mechanisms Underlying Developmental Shifts in Plant Evolution

OPEN ACCESS

Edited by:

Elena M. Kramer,
Harvard University, United States

Reviewed by:

Lachezar A. Nikolov,
University of California, Los Angeles,
United States

Madelaine Elisabeth Bartlett,
University of Massachusetts Amherst,
United States

*Correspondence:

Natalia Pabón-Mora
lucia.pabon@udea.edu.co

Specialty section:

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

Received: 31 March 2019

Accepted: 13 May 2019

Published: 04 June 2019

Citation:

Pabón-Mora N, Di Stilio VS and
Becker A (2019) Editorial: Genetic
Regulatory Mechanisms Underlying
Developmental Shifts in Plant
Evolution. *Front. Plant Sci.* 10:710.
doi: 10.3389/fpls.2019.00710

The origin and evolution of land plants was accompanied by major macroevolutionary changes, including profound shifts in reproductive processes. Originally, plants were heavily dependent on water for gamete transfer, while later on, male gametes were dispersed by wind or animals within highly reduced gametophytes. Over time, female gametophyte development was internalized and sperm motility was lost, requiring a sperm delivery system reaching deep into the maternal tissues inside a megasporangium that evolved into the ovule in seed plants. Moreover, seed plant embryos became protected within the seed, which contains a nutritional start-up package for the next generation. In flowering plants, both gamete- and seed-dispersal mechanisms diversified. A carpel surrounded the ovules requiring fertilization to occur internally and endosperm developed with the embryo as a result of double fertilization, presumably reducing the risk of allocating resources to inviable seeds. Flowers developed unlimited displays for pollination, involving mostly variations in perianth presentation that include elaborate coloration and symmetry changes. Finally, as carpels mature into fruits, they showcase a wide array of forms that guarantee optimal seed dispersal.

Elaboration of the vegetative phase of land plants also played a major role in their evolution and habitat adaptation. For instance, fixed multicellularity in the diploid phase of the life cycle (i.e., embryo formation) marks the transition to land. Embryo patterning required the diversification of cell types, the polarization of stem cell niches in the diploid axis and the occurrence of lateral organs in response to auxin peaks to form leaf primordia. In the sporophyte, leaves develop from the Shoot Apical Meristem (SAM) achieving extreme morphological variation. Roots developing from a Root Apical Meristem (RAM) provided anchoring to the substrate and access to nutrients and water becoming another key innovation of the independent sporophyte. In parallel, the exposure to new environments triggered new symbiotic interactions.

Central genes that regulate each of these processes act together in larger gene regulatory modules and networks, and numerous components and interactions are known in the model plant *Arabidopsis thaliana* (thale cress). However, how the gene regulatory modules have arisen over evolutionary time to enable these evolutionary transitions remains an open question at the core of plant evolutionary developmental research studies. This special issue collects articles contributing to this question from different perspectives and systems.

ORIGINS OF COMPLEXITY: THEORETICAL APPROACHES AND THE STUDY OF GENE EVOLUTION OVER DEEP PHYLOGENETIC TIMESCALES

Complexity in plant body patterning has been addressed from both morphological and genetic standpoints in this issue. The article by Benítez et al. aims to understand the mechanisms responsible for the major body plan transitions in green algae belonging to different clades and the monophyletic land plants. They introduce the concept of dynamical patterning modules (DPMs) which are defined as sets of conserved gene products and molecular networks, in conjunction with the physical morphogenetic and patterning processes where they function. They note that critical DPMs for the evolution of plants are cell-to-cell adhesion, placement of the cell wall, cell differentiation, and cell polarity. Analyzing data from broad phyletic comparisons, they conclude that the same DPM patterns have emerged many times independently, and that unicellular plant ancestors already possessed most molecular mechanisms later co-opted by multicellular plants, regardless of whether the diploid or haploid life phase is the dominant one. They further show that plasmodesmata were critical for the evolution of complex multicellularity in plants.

Yruela et al. highlight the evolution of protein ductility (intrinsic disorder) in unicellular and multicellular organisms. Investigating the occurrence of protein ductility associated with gene duplication led them to conclude that it increases in concert with organismic complexity. Moreover, the distribution of intrinsically disordered proteins and residues is not random but can be correlated with chromosomal rearrangement during evolution.

Complexity can also be studied in terms of gene regulation over developmental or evolutionary timescales. Bräutigam and Cronk examined the role of DNA methylation in gene regulation to fine-tune developmental plasticity. DNA methylation plays a role in enlarging the potential of phenotypic expression and the authors argue for the emergence of the novel research field of “epi-evo-devo.” Complexity can also be modulated by changes in developmental timing; Buendia-Monreal and Gillmor set up a framework in which alterations in the timing of developmental programs during evolution (also known as heterochrony) led to changes in the diversification of key innovations such as leaves, roots, flowers and fruits.

Synapomorphies in plant evolution can also be studied from the evolutionary perspective of gene lineages controlling developmental, morphological and physiological changes. For example, *AINTEGUMENTA* genes encoding AP2-type transcription factors play multiple functions in plant development including the maintenance of stem cell niches, embryo patterning, lateral organ formation, and fatty acid metabolism. Dipp-Álvarez and Cruz-Ramírez provide a comprehensive evolutionary framework for the *ANT* gene lineage across streptophytes, where *preANT-like* genes are present in the ancestor of embryophytes and a gene duplication occurs in land plants resulting

in the *basalANT* and the *euANT* lineages. Possible roles facilitated by these transcription factors during the transition to dry environments include enhanced tolerance to desiccation, and the establishment and maintenance of multicellularity.

The above-mentioned articles contribute to a theoretical and experimental foundation for assessing how major novelties may have arisen during plant evolution, and how potential fine-tuning of the genetic components responsible for a trait can lead to considerable variation in plant body plans.

ORIGIN AND PATTERNING OF THE GAMETOPHYTE AND SPOROPHYTE IN EARLY-DIVERGING LAND PLANTS

Three manuscripts focused on early-diverging lineages, including non-vascular bryophytes and vascular lycophytes. Flores-Sandoval et al. provide an overview of transcription factors controlling developmental transitions during the life cycle of the model liverwort *Marchantia polymorpha*. The authors explore differentially expressed genes during specific developmental time points, resulting in sets of genes acting exclusively in either gametophyte or sporophyte, in the reproductive transition or in antheridia or archegonia development. Their analyses reveals auxin co-expression groups present in liverworts and mosses that are not dependent on the single class-C Auxin Response Factor (ARF), which in other plant groups is directly involved in auxin responses. Moreover, their data confirm the participation of *MpARF3* as a negative regulator of reproductive transition.

The article by Grosche et al. shows that the common symbiotic pathway required for the mutualistic interactions between plants and mycorrhiza and the downstream GRAS-domain transcription factors important for its establishment were already present in the land plant ancestor. Using phylogenetic reconstructions, they showed moss lineage specific gene losses and expansions as well as absence of symbiotic GRAS TFs in algae. The ancestral presence of symbiotic GRAS TFs may have therefore provided a platform for conquering the land by enhancing microbial interactions. This idea is supported by the fact that in mosses, gene losses in some lineages coincide with lack of arbuscular mycorrhiza symbiosis.

A different approach was taken by Augstein and Carlsbecker who were able to trace roots into two independent origins in lycophytes and euphyllophytes (which include ferns and seed plants). They review the anatomical diversity of roots, emphasizing stele patterning, the auxin pathways in the RAM in different taxa, the genetic mechanisms involved in stem cell niche maintenance and the root cap control. Most data gathered so far comes from *Arabidopsis*, other Brassicales, the ferns *Azolla filiculoides* and *Ceratopteris richardii* and the lycophyte *Selaginella moellendorffii*. However, the authors emphasize that functional tools for comparative analyses are needed in lycophytes and ferns in order to establish the conservatism of such networks across vascular plants.

EVOLUTION OF THE MEGASPORANGIUM (OVULE) AND MEGAGAMETOPHYTE IN SEED PLANTS

Gymnosperms showcase unique reproductive features that include the nourishment of the embryo directly by the female gametophyte and the nucellus, which play the equivalent role of the endosperm in angiosperms. Moyano et al. describe the dismantling of the megagametophyte during germination of *Araucaria angustifolia* seeds, focusing on the mechanisms activating programmed cell death (PCD) and the availability of proteins, starch and lipid bodies for the developing embryo.

Ovule integuments are another distinctive feature distinguishing gymnosperms, which have one, from angiosperms, which have typically two. Integuments protect the ovules and seeds, define a route for pollen entry and contribute to seed hydration and dormancy. In an effort to assess the molecular basis of integument identity Arnault et al. investigate the expression of integument genes in the early-diverging angiosperm *Amborella trichopoda* and conclude that *YABBY*, *KANADI*, and *HD-ZIP III* transcription factors have conserved expression patterns between *Amborella* and *Arabidopsis*. Their data contribute to the reconstruction of molecular mechanisms for integument identity during the evolution of angiosperms.

Akhter et al. explore the evolution of gene lineages that exclusively expanded in gymnosperms compared to angiosperms employing RNA sequencing. They focus on the *TM3-like* MADS box genes, from which *DAL19* has been previously identified as being specifically upregulated in cone-setting shoots. They show the importance of previously unrecognized, and sometimes mutually exclusive, mRNA splice variants in *Picea abies*. They also highlight isoforms that are differentially expressed in male and female cone meristems, as well as in vegetative meristems. From their work, we derive that splice variants are in fact another source of variation contributing to functional evolution, working in parallel with gene duplication.

EVOLUTION OF MOLECULAR MECHANISMS UNDERLYING FLOWERING

Gene duplication is a known source for variation and functional diversification, triggering major evolutionary shifts. Flowering is one of the most important transitions leading to angiosperms; during this process, the SAM becomes an inflorescence meristem that in turn develops floral meristems in its flanks. Lee et al. exemplify the role of gene duplication in one of the key players in the transition to flowering, *FLOWERING LOCUS C (FLC)* in *Boechera stricta* (Brassicaceae), resulting in the acquisition of unique roles in the different paralogs. While one of the paralogs plays a conserved role delaying flowering, the other has lost its flowering function altogether. The authors uncover independent mutations that change the species phenology and provide evidence for heritable variations in vernalization requirements and flowering time via *FLC* in Brassicales.

A mini review by Monniaux and Vandebussche discusses perianth evolution. They propose that the perianth is formed

in the outer floral whorls to maintain the stamen and carpel identity gene's expression exclusively in the center of the flower. They evaluate negative regulators of B and/or C-class genes, especially of the *APETALA2 (AP2)* type and highlight the need for a broader comparative framework including early-diverging angiosperms with and without perianth. Exploring the molecular mechanism underlying floral organ identity Galimba et al. investigate the genetic redeployment of B-class genes in apetalous *Thalictrum* (Ranunculaceae). Ranunculaceae petals have been lost repeatedly, presumably in conjunction with the loss of the petal-specific *AP3-III* paralog. The authors present evidence for partial redundancy for stamen identity in the remaining paralogs and a role for these genes in the ectopic petaloidy of sepals, while proposing a novel mechanism of dominant-negative regulation of B-class genes by a truncated *AP3-II* ortholog.

Damerval et al. explore the genetic underpinnings of floral symmetry changes affecting floral display in Proteaceae, an early diverging eudicot family. They find that in *Grevillea juniperina*, adnation of stamens to tepals and asymmetrical growth of the single carpel contribute to the establishment of floral bilateral symmetry. An annotated floral transcriptome for *G. juniperina* is also presented, with an emphasis on floral MADS-box genes and TCP Class I and Class II gene expression patterns, the latter known to control floral bilateral symmetry in core eudicots.

Contributing to a more conceptual understanding of floral symmetry control, Sengupta and Hileman discuss the idea of direct transcriptional autoregulation (DTA) of *CYCLOIDEA (CYC)* genes. They present *in silico* predictions and experimental evidence for DTA in flower symmetry evolution and propose that *CYC* autoregulation may have evolved via *de novo* mutations and could have played a key role in the origin of monosymmetric flowers in the Lamiales.

EVOLUTION OF THE GENETIC NETWORK CONTROLLING FRUIT DEVELOPMENT

A fruit genetic network is well established in *Arabidopsis*, but comparative analyses of the key players controlling histogenesis during fruit development outside Brassicaceae is still scarce. Zumajo-Cardona et al. present the evolution of the *REPLUMLESS (RPL)* gene lineage, focusing on the expression patterns of *RPL* homologs in Papaveraceae (a basal eudicot). *Arabidopsis RPL* controls the identity of the replum, a medial persistent fruit layer unique to Brassicaceae fruits. In contrast, *RPL* homologs control fruit shedding in rice, whereas in poppy they are broadly expressed during flower development and become restricted to the dehiscence zone during fruit maturation, suggesting shifting roles of *RPL* genes during angiosperm diversification.

Maheepala et al. present a characterization of Solanaceae *FRUITFULL (FUL)* genes. *FUL* is responsible for fruit wall proliferation and for limiting the dehiscence zone in the *Arabidopsis* silique. While *FUL* homologs play the same roles in dry-fruited Solanaceae, they have taken on new roles in fleshy fruit development where they regulate aspects of the ripening processes, such as pigment accumulation. The

authors show that Solanaceae have four *FUL* paralogs, some originating as a result of a whole genome multiplication event, others by tandem duplication and one clade even undergoing pseudogenization. While some Solanaceae *FUL* clades appear to have acquired novel functions in fleshy fruit development, the molecular mechanisms underlying the *FUL* function shifts require additional analyses.

In summary, this research topic explores the genetic mechanisms controlling key developmental transitions, both vegetative and reproductive, during plant evolution. It includes original contributions on a variety of scales and processes: from factors contributing to multicellularity, to body plan complexity, developmental plasticity, heterochrony, tolerance to desiccation, gametophyte to sporophyte transition, establishment of embryo polarity and elaboration of the apical and root meristems and symbiotic interactions in early diverging land plants. Recognizing that reproductive shifts have also occurred in more recent phylogenetic scales, this collection also includes manuscripts focusing on the control of flowering, the development of ovules and perianth, floral symmetry and display and the elaboration of fruits from carpels aimed at dispersing the next generation. We hope that such comprehensive overview will be inspiring and will motivate additional efforts in the scientific community to continue to explore these processes holistically across land plants.

AUTHOR CONTRIBUTIONS

NP-M wrote the first draft of this manuscript, VSD and AB revised it and completed the text. All authors made direct intellectual contribution to the work and approved it for publication.

FUNDING

NP-M thanks funding from Universidad de Antioquia Convocatoria Programáticas 2017-16302 COLCIENCIAS 808 retos de país grant number 111580863819. VSD acknowledges funding from The Fred C. Gloeckner Foundation, Inc. AB is grateful for funding from German Research Foundation (DFG), grant numbers BE2547/12-1,2 and 14-1.

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 Pabón-Mora, Di Stilio and Becker. 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.



Dynamical Patterning Modules, Biogeneric Materials, and the Evolution of Multicellular Plants

Mariana Benítez¹, Valeria Hernández-Hernández², Stuart A. Newman³ and Karl J. Niklas^{4*}

¹ Centro de Ciencias de la Complejidad – Instituto de Ecología, Universidad Nacional Autónoma de México, Mexico City, Mexico, ² Laboratoire de Reproduction et Développement des Plantes, Université de Lyon, École Normale Supérieure de Lyon, Université Claude Bernard Lyon 1, Centre National de la Recherche Scientifique, Institut National de la Recherche Agronomique, Lyon, France, ³ Department of Cell Biology and Anatomy, New York Medical College, Valhalla, NY, United States, ⁴ Plant Biology Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, United States

OPEN ACCESS

Edited by:

Verónica S. Di Stilio,
University of Washington,
United States

Reviewed by:

Daniel H. Chitwood,
Michigan State University,
United States
Olivier Hamant,
École Normale Supérieure de Lyon,
France

*Correspondence:

Karl J. Niklas
kjn2@cornell.edu

Specialty section:

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

Received: 13 April 2018

Accepted: 04 June 2018

Published: 16 July 2018

Citation:

Benítez M, Hernández-Hernández V,
Newman SA and Niklas KJ (2018)
Dynamical Patterning Modules,
Biogeneric Materials,
and the Evolution of Multicellular
Plants. *Front. Plant Sci.* 9:871.
doi: 10.3389/fpls.2018.00871

Comparative analyses of developmental processes across a broad spectrum of organisms are required to fully understand the mechanisms responsible for the major evolutionary transitions among eukaryotic photosynthetic lineages (defined here as the polyphyletic algae and the monophyletic land plants). The concepts of dynamical patterning modules (DPMs) and biogeneric materials provide a framework for studying developmental processes in the context of such comparative analyses. In the context of multicellularity, DPMs are defined as sets of conserved gene products and molecular networks, in conjunction with the physical morphogenetic and patterning processes they mobilize. A biogeneric material is defined as mesoscale matter with predictable morphogenetic capabilities that arise from complex cellular conglomerates. Using these concepts, we outline some of the main events and transitions in plant evolution, and describe the DPMs and biogeneric properties associated with and responsible for these transitions. We identify four primary DPMs that played critical roles in the evolution of multicellularity (i.e., the DPMs responsible for cell-to-cell adhesion, identifying the future cell wall, cell differentiation, and cell polarity). Three important conclusions emerge from a broad phyletic comparison: (1) DPMs have been achieved in different ways, even within the same clade (e.g., phycoplastic cell division in the Chlorophyta and phragmoplastic cell division in the Streptophyta), (2) DPMs had their origins in the co-option of molecular species present in the unicellular ancestors of multicellular plants, and (3) symplastic transport mediated by intercellular connections, particularly plasmodesmata, was critical for the evolution of complex multicellularity in plants.

Keywords: plant evolution, plasmodesmata, algal evolution, convergent evolution, dynamical patterning modules

INTRODUCTION

The goal of this paper is to review the evolution of the multicellularity plant body plan within the conceptual framework of dynamic patterning modules (DPMs; Newman and Bhat, 2009; Newman et al., 2009; Newman, 2011), which provides a means of integrating physical and molecular-genetic aspects of developmental mechanisms. We have reviewed this topic previously (Hernández-Hernández et al., 2012; Niklas and Newman, 2013; Niklas, 2014). However, our focus here is on

the DPMs involved in the establishment of body plan polarity and cell-tissue differentiation. As in our previous treatments of the topic, a broad comparative approach is adopted here because multicellularity has evolved multiple times among the various eukaryotic photosynthetic lineages (**Figure 1**). The exact number of times it has evolved in large part depends on how multicellularity is defined. If multicellularity is regarded as any transient or permanent aggregation of cells, it is estimated to have evolved independently at least 25 times (Grosberg and Strathmann, 2007). If more rigorous criteria are applied, as for example the requirement for intercellular communication and cooperation, multicellularity has evolved multiple times in the Actinobacteria, Myxobacteria, and Cyanobacteria, at least three times in the fungi (chytrids, ascomycetes, and basidiomycetes), six times among the polyphyletic algae (twice each in the red, brown, and green algae), but only once in the Animalia (Niklas and Newman, 2013; Niklas, 2014).

Regardless of how multicellularity is defined or how many times it evolved, the repeated independent evolution of multicellularity evokes many important but as yet unresolved questions. For example, are the developmental and morphological motifs involved in the transformation of unicellular organisms into multicellular ones adaptations to the exigencies of life, the result of weak selection pressures, or the predictable consequences of physico-genetic laws and processes? Do sets of “master genes” for multicellularity exist among most or all eukaryotic clades? Indeed, given its ubiquity among pro- and eukaryotic lineages, has multicellularity truly

evolved independently among so many kinds of bacteria, fungi, algae, land plants, and animals, given the fact that all eukaryotes ultimately shared a last common ancestor?

The application of the framework of DPMs is particularly useful to address this last question because similar if not identical phenotypes can be achieved by the developmental mobilization of very dissimilar molecular systems or processes and because natural selection acts at the level of the phenotype and not at the level of the mechanisms that give rise to it. This dictum has been formalized by Newman and Bhat (2009), Newman et al. (2009), and Newman (2011) who have conceptualized the development and evolution of multicellular animals in the framework of DPMs each of which involves one or more sets of shared gene networks, their products, and the physical processes that relate to various types of matter. The importance of many of the physical processes involved in DPMs such as adhesion, cohesion, diffusion, activator-inhibitor dynamics, and viscoelasticity have long been recognized as important in development. Moreover, experimental research continues to demonstrate that the mechanical environment experienced by individual cells, tissues, and organs can alter gene expression patterns and thus cell fate specification (e.g., Swift et al., 2013).

Considering development and its evolution in terms of the DPM framework highlights the fact that the morphological motifs that are produced by physical processes evoked by specific molecules and pathways constitute a “pattern language” for configuring the basic body plans of multicellular animals and plants (Newman and Bhat, 2009; Newman et al., 2009; Hernández-Hernández et al., 2012; Niklas, 2014). These processes include mechanical forces resulting from the geometrical arrangements of mesoscale materials, irreversibility, the properties of network topologies and organization, and symmetry breaking. Importantly, many of the physical processes associated with DPMs are “generic” in that they are causally similar to the physical processes affecting the behavior of inorganic materials (Niklas, 1992; Niklas and Spatz, 2012). This congruence between the animate and inanimate world facilitated the rapid evolution of stereotypical generic morphologies once multicellularity was achieved in phylogenetically different groups of organisms, because there is ample evidence that some DPMs originated by means of the co-option of genes or gene regulatory networks (GRNs) present in ancestral unicellular organisms (Newman and Bhat, 2009; Newman et al., 2009).

Numerous examples of analogous DPMs operating across a broad spectrum of eukaryotic organisms can be given because of the many fundamental similarities existing among all eukaryotic cells (Wayne, 2009). For example, molecular pathways for the control of cell shape and polarity that evolved in unicellular organisms were mobilized by the novel protein Wnt in multicellular animals to mediate, via respective DPMs, lumen formation, and tissue elongation via convergent extension (reviewed in Newman, 2016b). Likewise, all eukaryotic cells have the capacity to produce extracellular (ergastic) polysaccharides and structural glycoproteins capable of self-assembly to create extracellular matrices containing interpenetrating polymeric networks of hydroxyproline-rich glycoproteins, e.g., collagen in animals and the extensin

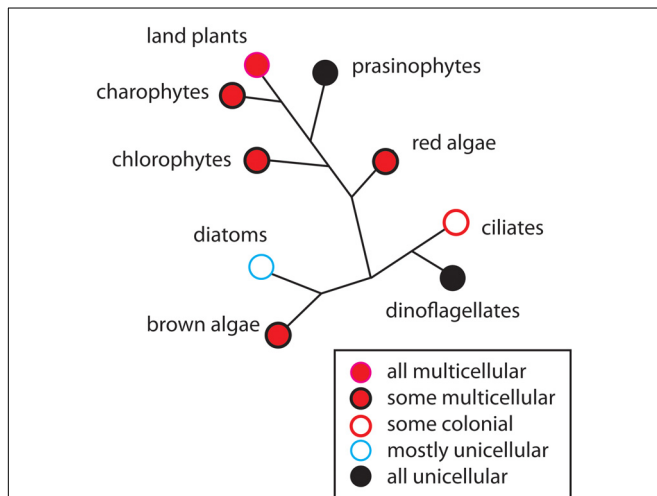


FIGURE 1 | The occurrence of multicellularity shown on a highly redacted and unrooted phylogenetic diagram of the major groups of photosynthetic eukaryotes. Although some groups are entirely unicellular or multicellular (e.g., prasinophytes and the land plants, respectively), most contain a mixture of body plans such as the unicellular and colonial body plans (e.g., diatoms), or a mixture of the unicellular, colonial, and multicellular body plans (e.g., brown algae). In general, early-divergent persistent lineages are dominated by unicellular species (e.g., prasinophytes in the green algal clade), whereas later-divergent lineages contain a mixture of body plans (e.g., chlorophytes and charophytes). Species-rich, late-divergent persistent lineages tend to be exclusively multicellular (e.g., the land plants and metazoans).

superfamily in numerous algal lineages and in the embryophytes (Ferris et al., 2001; Stanley et al., 2005) (**Figure 2**). These proteins manifest marked peptide periodicity, can form flexible rod-like molecules with repeat motifs (dominated by hydroxyproline) in helical configurations with arabinosyl/galactosyl side chains. This “superfamily” of intercellular adhesives operates among many unicellular organisms in gamete-to-gamete self-recognition and adhesion and the adhesion of cells to a substratum. It is very likely, therefore, that these “ancestral” adhesive capacities were co-opted to provide the cell-to-cell adhesives operating in many multicellular organisms, just as a wide array of microtubule-associated proteins in the algae, embryophytes, fungi, and metazoans (Gardiner, 2013) mediate related cell reshaping mechanisms utilized by DPMs in all these groups.

It is clear, however, that some of the DPMs operating in animals do not function in the various groups of algae, the land plants, and most fungi because of substantive differences among these lineages and clades (Meyerowitz, 2002; Newman and Niklas, 2018). Consider, for example, that the cells of animals are typically individually deformable and that during development they are free to move past one another in ways that permit differential adhesion, cortical tension, and other processes that permit the autonomous sorting and assembly of different tissues. In contrast, most plant and fungal cells possess a rigid cell wall that is firmly bound to the cell walls of adjoining cells. Likewise, plant signaling molecules acting as transcriptional modulators and determinants of tissue and cell fate can act intercellularly as well as intracellularly (see Cui et al., 2007; Urbanus et al., 2010; Garrett et al., 2012). This capacity, which is rare albeit not unknown in animal systems (Prochiantz, 2011), blurs the functional distinction of the GRNs affecting multi- versus single-cell differentiation.

Further, plant cell polarity involves the participation of PIN and PAN1 proteins in auxin polar and lateral transport and the regulation of metabolic fluxes by means of plasmodesmata. In contrast, animal cell polarity involves the participation of

integrin, cadherin, and PAR or CDC42 proteins (Geldner, 2009; Dettmer and Friml, 2011; Zhang et al., 2012). Further, cell, tissue, and organ polarity within the multicellular plant body is maintained by a complex phytohormone transport system that involves the differential and sometimes transient positioning of auxin transporters proteins (Dettmer and Friml, 2011; Zhang et al., 2012), the establishment of mechanical heterogeneities within the apoplastic infrastructure (Kutschera and Niklas, 2007; Geldner, 2009; Peaucelle et al., 2015; Galletti et al., 2016; Majda et al., 2017), and the regulation of metabolic fluxes by means of plasmodesmata. Although analogies can be drawn between the establishment and maintenance of cellular and tissue polar domains in animals and plants, the mechanisms by which polarity is achieved are very different. For example, tight junctions between the apical and basolateral plasma membrane domains in animal epithelial cells provide barriers preventing the intramembrane diffusion of proteins and other macromolecules (Shin et al., 2006), whereas in plants a variety of phenolic compounds are used to maintain tissue polarity domains (Allassimone et al., 2010). As a final example of the differences between DPMs among the eukaryotic lineages, consider the manner in which cell wall materials are delivered and deposited during cell division. The mechanics of this developmental process differs substantively among the desmids and among different filamentous ascomycetes (Hall et al., 2008; Seiler and Justa-Schuch, 2010). It even differs within the monophyletic Chlorobionta, i.e., phycoplasmic cell division in the Chlorophyta and phragmoplasmic cell division in the Streptophyta (see Graham et al., 2009; Niklas, 2014) (**Figure 3**).

Focusing on the distinctive physico-genetic morphogenetic modalities of plants, Hernández-Hernández et al. (2012) identified six DPMs involved in critical embryophyte developmental processes. These DPMs are (1) the production of intercellular adhesives (ADH), (2) the manner in which the future cell wall is formed and oriented (FCW), (3) the establishment of intercellular communication and spatial-dependent patterns of differentiation (DIF), (4) the establishment of axial and lateral polarity (POL), (5) the formation of lateral appendages or “buds” (BUD), and (6) the formation of lateral, leaf-like structures (LLS). Hernández-Hernández et al. (2012) discussed all six of these DPMs with an emphasis on the first four (i.e., ADH, FCW, DIF, and POL), because cell division, cell-to-cell adhesion, intercellular communication, and polarity are essential for achieving simple multicellularity across all clades and because these four DPMs operate in a pairwise manner in many multicellular algae and fungi as well as in the land plants (**Figure 4**). Here, we emphasize DIF and POL because these are essential for achieving complex multicellularity (define here as “the condition in which some cells are not in direct contact with the external environment”), and present new evidence that the evolution of plasmodesmata played a critical role in the evolution of cell, tissue, and organ differentiation and polarity. We also identify the characteristic molecules and molecular networks, and when possible, the physical processes they mobilize for each of the four key modules.

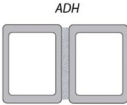
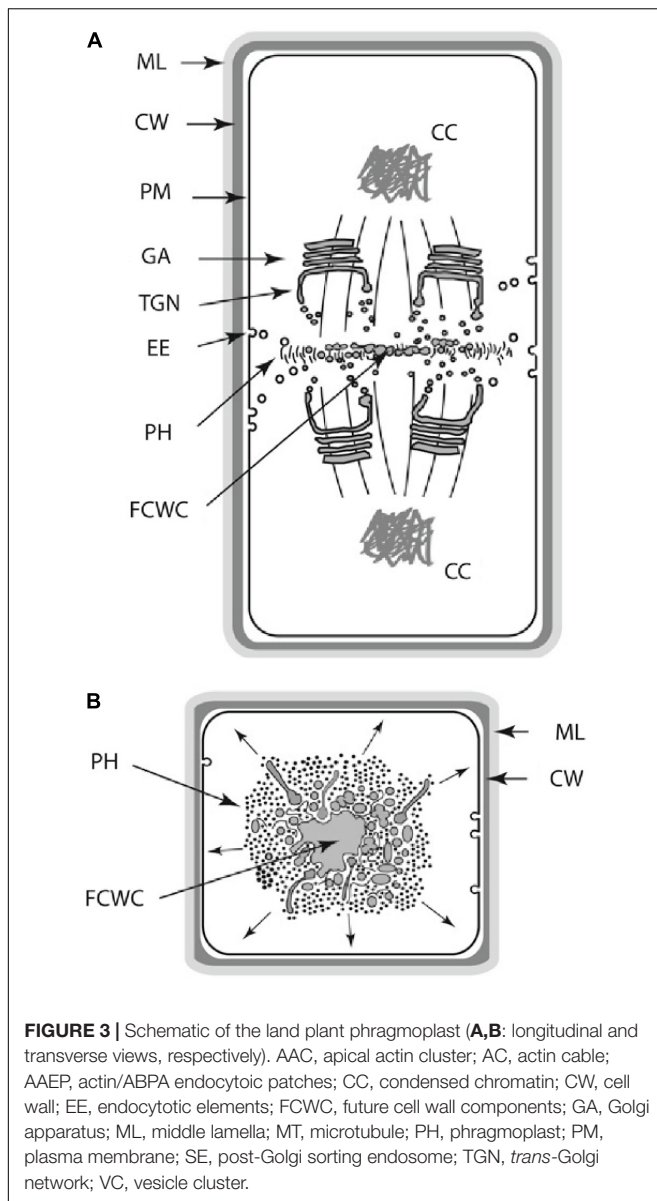
Lineage	Intercellular Adhesives	
Cyanobacteria	glycoproteins, lipopolysaccharides	Adhesion
Fungi	glycoprotein-based “glues”	
Rhodophytes	sulfated galactan polymers	
Phaeophytes	β -D-mannuronic and α -L-guluronic acid polymers	
Chlorophytes	hydroxyproline-rich glycoproteins	
Embryophytes	Ca-rhamnogalacturonic-dominated pectins	
Metazoa	type-1 transmembrane cadherin proteins	

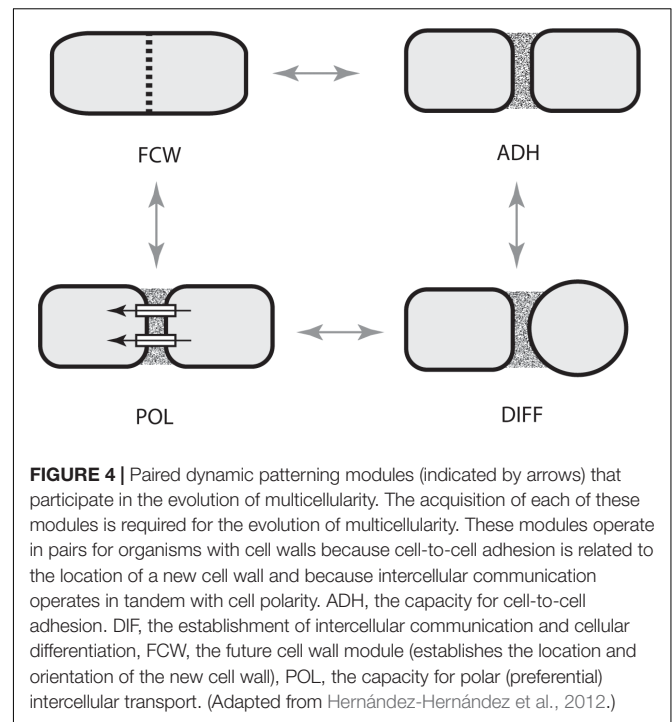
FIGURE 2 | The different adhesives utilized by the cell-to-cell adhesive (ADH) dynamic patterning module among some of the major multicellular lineages.



BIOGENERIC MATERIALS AND DYNAMICAL PATTERNING MODULES (DPMs)

Before proceeding to an exploration of the empirical evidence for the concepts that will be pursued, it is important to establish clear definitions for what is meant by biogeneric materials and DPMs, particularly since these concepts may not be familiar to some researchers and because clarity in definitions is essential for clarity in thinking.

Like all matter, living matter manifests inherent morphogenetic properties, and characteristic morphological motifs that are in part expressions of inviolable physical laws and principles (Newman, 2017; see also Niklas and Spatz, 2012; Niklas, 2017). This idea is familiar to those who study the physical



sciences as is the notion that the operation of physical laws and principles is sensitive to scale (Niklas, 1994). On the macroscale, phenomena such as large-scale climate and oceanic systems generate fluidic patterns of various but discrete and recognizable morphologies. On the microscale, atoms and small molecules can be arranged and rearranged to form discrete molecules with well-defined chemical and physical properties. Living matter operates at an intermediate or *mesoscale*. Non-living mesoscale materials are familiar as solids, which can be amorphous or crystalline, and liquids, which can form vortices and waves. Living matter exhibits many of these generic physical properties (Green and Batterman, 2017).

Although all living cells have mesoscale properties in common (e.g., their cell membranes and cytoplasm are rheologically similar), we focus here exclusively on multicellular matter. For example, all tissues behave as viscoelastic materials (i.e., they behave as a combination of a liquid and a solid). However, the extent to which a material manifests viscoelasticity depends on the presence and quantity of its rigid components (Niklas, 1989, 1992). With few exceptions (e.g., bone and cartilage), most animal tissues are highly viscoelastic owing to an absence of rigid cell walls. In contrast, all plant and fungal tissues behave as deformable cellular solids because of the presence of rigid cell wall solids (e.g., cellulose and chitin) (Niklas, 1989, 1992). Animal, plant, and fungal tissues have shared generic properties, but they differ in the degree to which they respond to physical stresses. Thus, the subunits – cells – of animal tissues can be independently mobile and rearrange with respect to one another, particularly during development, when the tissues are more liquid-like than they are in the mature organism, whereas, with the exception of intrusive

growth, plant and fungal cells do not typically change their neighbors.

The generic physical properties of living matter lend predictability to the forms it can assume during development. Like non-living liquids, liquid crystals, and mixtures thereof, developing animal tissues can form immiscible layers, interior spaces, and undergo elongation (Forgacs and Newman, 2005). Like non-living deformable solids (undergoing, for example, accretion or melting), developing plant and fungal tissues can bud or branch (Fleury, 1999; Niklas and Spatz, 2012).

The liquid vs. solid nature of living tissues does not arise from the same subunit-subunit interactions that endow non-living materials with these properties. Instead of Brownian motion and the electronic weak attractive interactions among the molecular subunits of non-biological liquids, the cells in animal tissues move non-randomly by cytoskeletally generated forces and remain cohesive despite their translocation via transmembrane homophilic attachment proteins. In plant and fungal tissues, instead of the charge-based or covalent bonds of the atomic or molecular subunits of non-biological solids, the cells are cemented together by Ca^{2+} -rhamnogalacturonic-rich pectins, or members of the extensin superfamily of hydroxyproline-rich glycoproteins (Cannon et al., 2008; Lampert et al., 2011). For these and other reasons, the various viscoelastic and deformable solid materials that constitute living tissues have been termed “biogeneric” matter, in recognition of the predictability of their morphogenetic behavior and outcomes afforded by their generic properties, and the fact that these generic properties are dependent on evolved biological, rather than purely physical, effects (Newman, 2016a).

Another set of biogeneric properties that characterize living tissues, superimposed upon their identity as predominantly viscoelastic or non-deformable solid materials, is their *excitability*, that is, the ability to store energy and release it upon stimulation (Levine and Ben-Jacob, 2004; Sinha and Sridhar, 2015). Mechanically, chemically, and electrically excitable materials are not unknown in the non-living world (exemplified by loaded mousetraps, forest fires, and tunnel diodes) but they are uncommon. Multicellular systems are inevitably excitable, because their cellular subunits are biochemically, mechanically, and electrically active, the storage and controlled utilization of energy is intrinsic to all life (Lund, 1947).

During the development of multicellular organisms, communication among the cellular subunits can induce the spatiotemporal mobilization of mechanical, chemical, and electrical energy, leading to cellular pattern formation and morphogenesis. In animal embryos and organ primordia, this communication is generally short-range via extracellularly diffusible morphogens. However, mechanical and electrical fields can achieve nearly instantaneous long-range communication. In developing and remodeling plants and fungi, communication can be both intra- and intercellular and short and long-range. Like the biogeneric rheological and solid-state properties of animal, plant, and fungal tissues, the phenomena of excitability give rise to predictable morphological motifs – repetitive or fractal arrangements of ridges, appendages, venation patterns and cell types.

Dynamical patterning modules, defined above, are intrinsic to this “physico-genetic” account of the origin and evolution of multicellular life-forms (Newman, 2012). The DPM concept recognizes that the physical forces that shape tissues cannot be considered independently of the actual materials (cell collectivities and their molecules) that they act on. The activity of DPMs can be regulated in a given multicellular material (e.g., that characterizing the phylum to which a species belongs), leading to developmental transitions and phenotypic differences between members of a phylum. Insofar as the materials have biogeneric properties (as is the case with animal tissues, and to a large extent with plant and fungal tissues), specific DPMs will promote morphological outcomes familiar from the physics of non-living matter. In other cases, DPMs will mobilize physical forces to produce outcomes peculiar to varieties of living matter. Without exception, however, physics and genetics act together to effect morphological development.

The following sections will update earlier descriptions of DPMs in plant systems (Hernández-Hernández et al., 2012; Benítez and Hejatkó, 2013; Niklas and Newman, 2013; Niklas et al., 2013; Niklas, 2014; Mora Van Cauwelaert et al., 2015) and attempt to assign evolutionary roles to them.

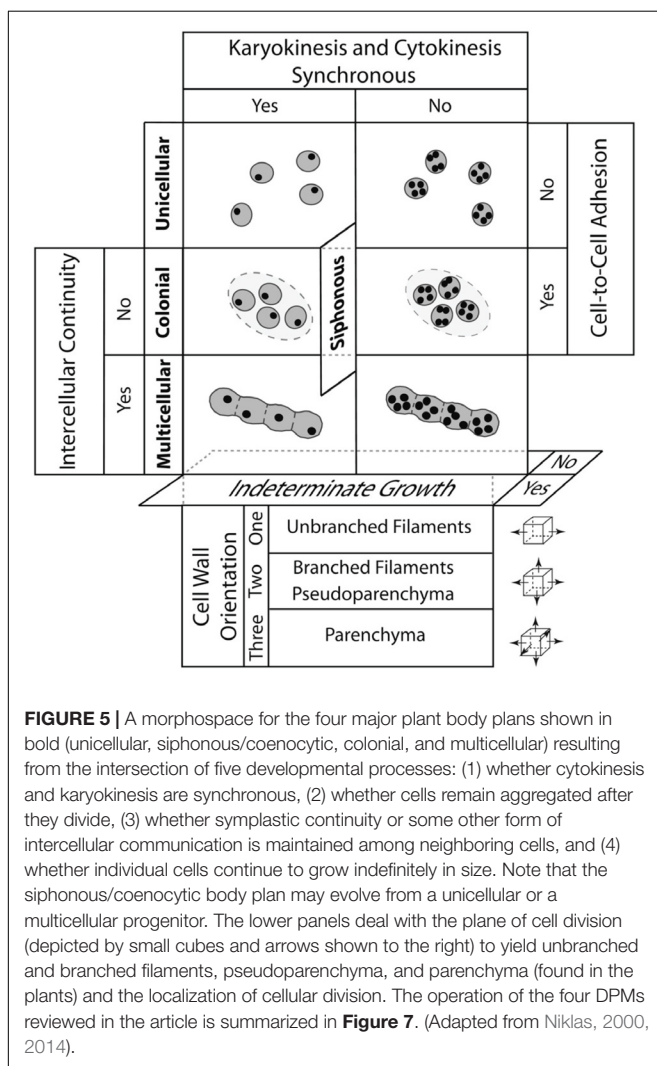
EVOLUTIONARY TRANSITIONS IN EUKARYOTIC PHOTOSYNTHETIC LINEAGES

The fossil record and contemporary molecular phylogenetic analyses indicate that the three major algal clades in which multicellularity evolved (i.e., the Streptophytes, Rhodophytes, and Chlorobionta) had independent evolutionary origins because of primary and secondary endosymbiotic events (reviewed by Kutschera and Niklas, 2005, 2008). Consequently, these three clades in tandem with the evolution of the land plants (from a green algal ancestor) can be viewed as independent “evolutionary experiments” that provide an opportunity to examine how the four DPMs (i.e., ADH, FCW, DIF, and POL) participated in achieving multicellularity in each case.

The significance of the four DPMs becomes apparent when they are placed in the context of a morphospace that identifies the major plant body plans and when their placement is juxtaposed with a series of phenotypic transformations predicted by multilevel selection theory for the evolutionary appearance of multicellularity regardless of the clade under consideration (Niklas, 2014). Following McGhee (1999), we define a morphospace as a depiction of all theoretically possible structural phenotypes within a specific group of organisms. The depiction is constructed using orthogonal axes, each of which represents a phenotypic character that has one or more character states, e.g., cellular aggregation: yes or no. The intersection of two or more such axes specifies a hypothetical or real phenotype defined by the variables or processes the participating and intersecting axes specify.

Niklas (2000, 2014) constructed such a morphospace for all photosynthetic eukaryotes using four characters, each of

which has two character states in the form of a question, i.e., (1) are cytokinesis and karyokinesis synchronous?; (2) do cells remain attached after cellular division?; (3) is symplastic or some other form of intercellular communication established and maintain among adjoining cells?; and (4) do individual cells continue to grow indefinitely in size? This simple morphospace identifies four plant body plans that can be either uninucleate or multinucleate, i.e., the unicellular, siphonous/coenocytic, colonial, and multicellular body plan (Figure 5). The different tissue constructs of the multicellular plant body can be identified also by adding a fifth axis that specifies the orientation of cell division with respect to the body axis. With the addition of this axis, three tissue constructs are identified, i.e., the unbranched filament (when cell division is restricted to one plane of reference), the branched filament and the pseudoparenchymatous tissue construct (when cell division occurs in two planes of reference and when branched filaments interweave, respectively), and the parenchymatous tissue construct (when cell division occurs in three planes of cell division).



The primary literature dealing with the algae (e.g., Graham et al., 2009) reveals that all four plant body plans (as well as the three tissue constructs) have evolved multiple times in the Stramenopiles, Rhodophytes, and Chlorobionta. This convergence reveals the evolutionary significance of the four DPMs that are the focus of our review, i.e., FCW, ADH, DIF, and POL (Figure 4). Specifically, ADH is required for the colonial and multicellular body plans; FCW is involved in whether cyto- and karyokinesis are synchronous and whether the tissue construction of a multicellular plant is filamentous (unbranched or branched) or parenchymatous, although how the FCW is determined remains problematic, even for the well studied land plants (Schaefer et al., 2017); and DIF and POL are required for intercellular cooperation and cellular specialization.

These four DPMs also help to identify evolutionary trends in the establishment of multicellularity predicted by multilevel selection theory (Folse and Roughgarden, 2010; Niklas and Newman, 2013; Niklas, 2014). This theory recognizes the unicellular organism as the ancestral state in each of the multicellular lineages or clades, and it identifies the colonial body plan as transitional to the multicellular body plan. Therefore, when multilevel selection theory is applied to the evolution of multicellularity, it identifies a “unicellular => colonial => multicellular” body plan transformation series (regardless of the type of organism) in which the participation of ADH, FCW, DIF, and POL are collectively required to establish and maintain a colonial body plan and to subsequently coordinate and specify the intercellular activities within an integrated multicellular body plan whose complexity exceeds simple dyadic interactions among conjoined cells, tissues, and organs.

The transformation series among the different genera and species of the volvocine algal lineage is consistent with the aforementioned multilevel selection theory’s predicted unicellular => colonial => multicellular transformation series (Bonner, 2000; Kirk, 2005; Herron and Michod, 2008; Niklas, 2014). The ancestral volvocine body plan undoubtedly possessed a unicellular organism that was morphologically and physiologically like *Chlamydomonas*. The transformation of this unicellular organism into a colonial organism is posited to have involved the modification of the ancestral cell wall into an extracellular adhesive matrix seen in the Tetrabaenaceae => Goniaceae => Volvocaceae transformation series (Kirk, 2005; see also Graham et al., 2009), which is consistent with what is known about the biochemistry of this ergastic material (Sumper and Hallmann, 1998). Subsequent evolutionary modifications exemplified by a hypothetical Goniaceae => Volvocaceae transformation series are predicted to have produced life-forms ranging from simple colonial aggregates (e.g., *Tetrabaena socialis*) to more colonies with asymmetric cell division, to multicellular organisms with a germ-soma division of labor (e.g., *Volvox carteri*) (Kirk, 2005; Herron and Michod, 2008). It is worth noting that in the case of multicellular volvocine algae, the cytoplasmic bridges that interconnect each cell to its neighbors have multiple functionalities. These bridges participate in the mechanics of a unique form of kinesin-driven inversion, and they provide avenues for the metabolic transport of nutrients to developing reproductive structures, called gonidia

(Hoops et al., 2000). In this sense, these bridges are analogous to land plant plasmodesmata, although their apertures are much wider than those of the latter (~200 nm in diameter; Green et al., 1981). Curiously, in some volvocines, these bridges are developmentally severed and thereby provide an interesting example of a multicellular-to-colonial transformation series.

DPMs AND BIOGENERIC PROPERTIES INVOLVED IN PLANT DEVELOPMENT AND EVOLUTIONARY TRANSITIONS

As noted previously, we proposed a set of DPMs associated with key plant developmental events and specified some of the physical and molecular components of these modules (Hernández-Hernández et al., 2012). After reviewing the phyletic distribution of the molecular elements of the DPMs, we hypothesized that these modules originated from the co-option of cell-molecular mechanisms related to single-cell functions in the unicellular ancestors of the major algal clades and the land plant lineages that mobilized, in the multicellular context, novel physical processes. One of our central conclusions is that once development is set into operation, much of it becomes self-organizing due to the mobilization of DPMs and biogeneric properties. This view contrasts with the hypothesis that land plant diversification resulted mainly from the expansion of particular gene families (e.g., Vergara-Silva et al., 2000; Zajímalová et al., 2010). Certainly, while these molecules are central for plant development and, most probably also for plant evolution, we argued that the notion that diversification of certain gene families or molecular classes can be the main cause of morphological evolution is insufficient. Additionally, we suggested that the combination of different DPMs at different places and developmental stages may help understand the generation of the basic features of the multicellular plant body plan. We argued further that plant development has evolved into processes that occur in a physical medium that is dynamic over large scales, utilizing inherently multicellular systems of multifunctional hormones/morphogens/transcription factors that are unrestricted by cell boundaries in many of their functions. Under such conditions, the origin and mechanisms behind plant extraordinary plasticity becomes less enigmatic (Hernández-Hernández et al., 2012).

The Spatially Dependent Differentiation (DIF) DPM and its Role in the Evolution of Multicellular Plants and Vascularization

As groups of cells adhered to each other, some physical constraints were imposed on the transport of nutrients and signaling molecules. Multicellular aggregates eventually evolved a division of labor (Niklas, 2000; Kirk, 2005; Knoll, 2011) that required cell fate specification mechanisms (Niklas et al., 2014), and the ability of cells to coordinate their metabolism, patterns of cell growth, and the activity of molecular networks. In almost every multicellular aggregate two possible mechanisms for the exchange of nutrients and signaling molecules exist:

indirect and direct transport (Beaumont, 2009). The first case requires some cells to secrete nutrients to the external environment and other cells to take them up (Beaumont, 2009). In contrast, for direct cell-to-cell transport, the presence of transmembrane connections is required (Niklas, 2000; Knoll, 2011). During the course of evolution, intercellular connections evolved independently in multicellular lineages to respond to the biophysical challenges that multicellularity imposed (Figure 6). For example, animals have gap junctions, whereas the cells of plants and fungi, respectively, developed plasmodesmata and septa pores, respectively (Bloemendal and Kück, 2013). Here, we discuss the role of plasmodesmata-mediated transport in the coordination of cell type specification during plant development and how this could have been a prerequisite for the transitions from unicellular to multicellular (and from non-vascular to vascular) plants. We will also briefly review the phylogenetic data concerning the evolution of plasmodesmatal structure.

Given that plant cells are surrounded by a rigid cell wall, they rely on the transport of signaling molecules for the establishment of cell type patterns. As we discussed in previous work (Hernández-Hernández et al., 2012), the spatially dependent differentiation (DIF) DPM is composed of the plant intercellular channels, called plasmodesmata, and the biogeneric properties they mobilize, viz., passive diffusion, lateral inhibition, and reaction-diffusion. Plants manifest a precise spatiotemporal control on the aperture (also called permeability) of plasmodesmata channels and can control the direction of plasmodesmata-mediated fluxes to create molecular concentration gradients (Sager and Lee, 2014). Using different techniques, Christensen et al. (2009) detected unidirectional transport of fluorescent probes from the basal epidermal cells into the apical cells of trichomes in the leaves of tobacco, *Nicotiana tabacum*. The probes were observed to move freely among trichome cells in both directions, but they were prevented from migrating in the opposite direction into subtending cells. Although the authors did not conclusively prove that this unidirectional flow depends on the aperture of plasmodesmata channels, they found that treatments with sodium azide, a metabolic inhibitor that alters plasmodesmata permeability, could reverse the direction of unidirectional flow from epidermal to trichome cells. A unidirectional transport of the photoconvertible dye Dendra2 was also observed in *Physcomitrella patens* (Kitagawa and Fujita, 2013), which indicated that this phenomenology is likely very ancient and thus of wide occurrence among the land plants. With the help of a controlled intercellular transport, plants can then modulate diffusion of signaling molecules in specific ways to generate or at least modulate patterns of cell type specification.

At the same time, the regulation of plasmodesmata permeability can generate morphogen gradients. Plasmodesmata aperture is regulated by the deposition and degradation of callose within the cell walls through which plasmodesmata pass (De Storme and Geelen, 2014). The turnover of callose is achieved by the participation of several families of proteins among which the GLUCAN SYNTHASE LIKE (GSL) proteins and β -glucanases, respectively, synthesize and degrade callose (Ruan et al., 2004; Guseman et al., 2010; De Storme and Geelen, 2014).

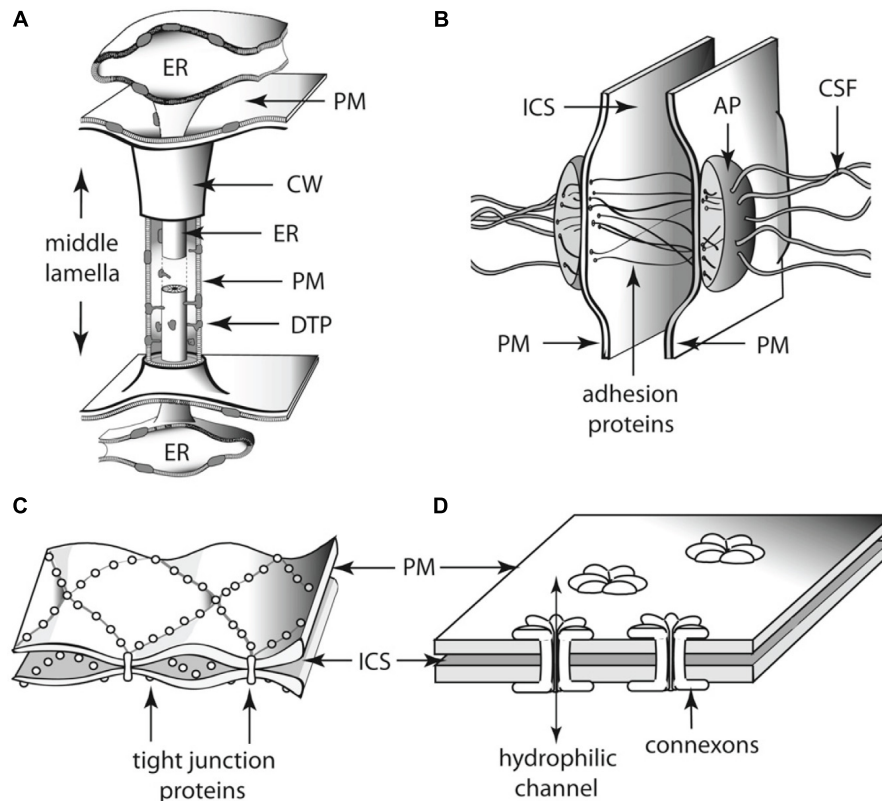


FIGURE 6 | Schematics of the diversity of intercellular communication among adjoining (A) land plant and (B–D) animal cells. Each of these cell-to-cell linkages participates in the establishment of cell polarity as well as physiological communication among adjoining cells. Thus, each represents an analogous evolutionary innovation for two of the essential features of multicellularity. (A) Plasmodesma. (B) Desmosome. (C) Tight junctions. (D) Gap junctions. AP, attachment plaque (plakoglobins); CSF, cytoskeletal filaments (keratin); CW, cell wall; DTP, desmotubular proteins; ER, endoplasmic reticulum; ICS, intercellular space; PM, plasma membrane.

Further, genetic and chemical experiments have correlated the amount of callose at plasmodesmata sites with the genetic expression of GSLs and β -glucanases, and the intercellular migration of molecules in several plant systems (Ruan et al., 2004; Guseman et al., 2010; Vatén et al., 2011; Benítez-Alfonso et al., 2013; Han et al., 2014). For example, in hypocotyls of *Arabidopsis* seedlings, it was demonstrated that the reduced callose deposition at plasmodesmata, resulting from an inducible knock down mutation of the *GLUCAN SYNTHASE LIKE 8* (*GSL8*) gene, had an enhanced diffusion of auxin (Han et al., 2014). Consequently, the loss of asymmetric auxin distribution prevented the differential cell elongation between the shaded and illuminated parts of the hypocotyl that is required for the phototropic response (Han et al., 2014). Based on these and other observations, Han et al. (2014) concluded that plasmodesmata closure is necessary to prevent auxin diffusion in *Arabidopsis* and to generate concentration gradients. In a similar way, it has been proposed that the main mechanism to establish auxin gradients in mosses such as *P. patens* is through plasmodesmata-mediated transport (Brunkard and Zambryski, 2017). Therefore, it seems likely that the regulation of plasmodesmata permeability has been key for land plants to establish concentration gradients of morphogens that coordinate developmental dynamics.

However, it is important to note that neither plasmodesmata nor multicellularity are required to achieve morphological complexity. This is evident from siphonous (coenocytic) algae such as the marine green alga *Caulerpa*. A recent intracellular transcriptomic atlas of this organism reveals that the acropetal transcript distribution conforms roughly to a transcription-to-translation pattern without the presence of internal cell walls (Ranjan et al., 2015; see also Menzel, 1996).

Cell type specification that depends on the intercellular transport of transcription factors is also accompanied by the closure of plasmodesmata to actuate a lateral inhibition mechanism. For example, the *chor* mutant of *Arabidopsis*, which encodes a putative *GSL8* protein, results in a significant increase in the number of stomatal lineage cells (Guseman et al., 2010). Further, in the epidermal cells of leaves, the expression of the *SPEECHLESS* (*SPCH*) transcription factor, which specifies the initiation of the stomatal lineage, is restricted to the meristemoid mother cells of stomata after asymmetric division (Pillitteri and Dong, 2013). The *gsl8* mutant has a lower amount of callose deposition resulting in the leakage of *SPCH* between epidermal cells that, in turn, results in abnormal stomata clusters (Guseman et al., 2010). By preventing the intercellular migration of *SPCH*, plasmodesmata inhibit the cells surrounding meristemoids to

differentiate into the stomata lineage and thus regulate the spacing of stomata in the epidermis of leaves. This demonstrates that the plasmodesmata aperture is necessary for the specification of cell identities by virtue of regulating lateral inhibition.

The non-cellular autonomous signaling mediated by symplasmic transport is a key mechanism to establish patterns of cell specification required for the development of vascular tissues. For example, in the root of *Arabidopsis* the transcription factor SHORT ROOT (SHR) moves from the stele into the cells within the quiescent center and the endodermis where it turns on the production of miRNA165/6 (Carlsbecker et al., 2010). The miRNA165/6 then moves back to the stele where it degrades the homeodomain leucine zipper PHABULOSA (PHB), which is necessary for the radial patterning of the xylem tissue and the pericycle (Carlsbecker et al., 2010). Mutations of the CALLOSE SYNTHASE GLUCAN LIKE 3/GLUCAN SYNTHASE LIKE 12 (CALS3/GSL12) gene, which the product of degrades callose, results in an increased callose deposition (Vatén et al., 2011). In these mutants, the signal of pSHR:SHR:GFP in the endodermis relative to that of the stele is decreased when compared with the wild type. This observation is consistent with the hypothesis that callose deposition prevents symplasmic transport (Vatén et al., 2011). Because of the downregulated symplastic transport, protoxylem cell identity was disrupted and metaxylem cells were ectopically expressed in the location of protoxylem cells (Vatén et al., 2011). In this manner, it is possible that plasmodesmata-mediated transport may have also driven the development of specialized cells and tissues by means of the spatiotemporal differential transport of nutrients.

Finally, it has become increasingly clear that the manner in which plasmodesmata are distributed within the multicellular plant body compartmentalizes this body into symplastic domains that can take on different functionalities by virtue of either sequestering aspects of metabolic activity, as for example during the dormancy of terminal tree buds (Tylewicz et al., 2018) or facilitating specific avenues of symplastic translocation, as for example the movement of mRNA within the phloem (Xoconostle-Cazares et al., 1999). When seen in this manner, the multicellular plant body plan is actually a continuous symplast incompletely partitioned by a continuous apoplast created by an infrastructure of perforated cell walls (Niklas and Kaplan, 1991).

Indeed, all the available evidence demonstrates the importance of plasmodesmata-mediated transport for plant development (Sager and Lee, 2014). Plasmodesmata seem to have appeared independently several times in the plant kingdom. Intercellular connections very similar to the plasmodesmata of land plants have been found in the multicellular species of the green, red, and brown algae (e.g., Cook et al., 1997, 1999; Raven, 1997). As in the case of the land plants, the plasmodesmata of the green alga *Bulbochaete hiloensis* are modulated during ontogeny in a manner that differentially limits intercellular transport and separates cellular domains into different functional identities (Fraser and Gunning, 1969; Kwiatkowska, 1999).

Some features of plasmodesmata seem to have evolved after the Chlorophyte–Streptophyte divergence. For example, there is some evidence that the encapsulation of the endoplasmic reticulum (ER) within the plasmodesmatal channel is unique to

the land plants. A close examination of plasmodesmata structure in the charophycean alga *Chara zeylanica* and in three putative early divergent bryophytes (the liverwort *Monoclea gottschei*, the hornwort *Notothylas orbicularis*, and the moss *Sphagnum fimbriatum*) reveals that in contrast to *C. zeylanica*, all three bryophytes have encased ER (Cook et al., 1997). The ER lumen serves as another pathway for intercellular transport making plasmodesmata transport more complex (Guenoune-Gelbart et al., 2008). The more complex plasmodesmata with internal ER of the land plants are present in some green algae, such as *Uronema* and *Aphanochaete* (Chlorophyceae) (Floyd et al., 1971; Stewart et al., 1973), and in some Laminariales brown algae (Marchant, 1976; Sideman and Scheirer, 1977). However, the movement of molecules through the lumen or the ER of these plasmodesmata has not been yet demonstrated for these algae. Based on these observations, it is reasonable to conclude that plasmodesmata lacking encased ER evolved first and that the encapsulation of ER is an evolutionarily derived feature that was present in the green algal ancestor of the land plants, well before bryophytes diverged (Lucas et al., 1993; Cook et al., 1997).

Given that the land plants are more complex than their algal ancestors because of the presence of specialized cell types and tissues for nutrient transport, we speculate that the increased complexity of multicellular plants is associated with the evolution of structurally complex encapsulated-ER plasmodesmata. This speculation emerges in part from a consideration of the limitations imposed by passive diffusion on the transport of metabolites and by the necessity of bypassing these limits as multicellularity resulted in larger and larger life-forms. Specifically, manipulation of Fick's second law of passive diffusion shows that the time it takes for the concentration of a non-electrolyte j initially absent from a cell's interior to reach one-half the concentration of j in the external ambient medium (denoted as $t_{0.5} - t_0$) is given by the formula

$$t_{0.5} - t_0 = \frac{V}{AP_j} \ln \frac{(c_0 - c_j)t_0}{(c_0 - c_j)t_{0.5}} = 0.693 \frac{V}{AP_j},$$

where V and A are the volume and the surface area of the cell, respectively, P_j is the permeability coefficient of j (a constant for any particular non-electrolyte), the expression $(c_0 - c_j)t_0$ is the initial difference between the external and internal concentrations of j at time zero, and the expression $(c_0 - c_j)t_{0.5}$ is the difference between the external and internal concentrations when the internal concentration of j reaches one-half that of the ambient medium (Niklas and Spatz, 2012). This formula shows that the time required for passive diffusion to provide essential metabolites to a cell increases in direct proportion to the volume of a cell. Beyond a certain surface area-to-volume limit, passive diffusion must be replaced by bulk flow, which is impossible within a unicellular non-aquatic organism. Consequently, the evolution of complex multicellularity requires intercellular bulk flow that necessitates some form of intercellular "porosity," e.g., phloem sieve plates. Likewise, intercellular transport systems require cell-type specialization, which has been shown to be positively correlated with genotypic and proteomic "complexity" (e.g., Niklas et al., 2014, 2018; Yruela et al., 2017).

Molecules that regulate plasmodesmata aperture and structure may have performed different functions in the ancestors of land plants. As previously noted, callose turnover is the main contributor to the regulating of plasmodesmata permeability. Although callose is widespread in the plant kingdom, its turnover regulated by plasmodesmata aperture has only been observed in the land plants (Scherp et al., 2001; Schuette et al., 2009). Thus, understanding the functionalities of plasmodesmata-localized proteins implicated in callose turnover could help elucidate the evolution of plasmodesmata structure and plasmodesmata-dependent transport. For example, glycosyl hydrolase 17 (GHL17) belongs to another family of proteins involved in callose degradation (Gaudioso-Pedraza and Benítez-Alfonso, 2014). A phylogenetic study using the sequences of GHL17 of fungi, algae, bryophytes, *Arabidopsis*, and monocots identifies a land plant specific clade characterized by plasmodesmata GHL17 localization (Gaudioso-Pedraza and Benítez-Alfonso, 2014). In contrast, the fungal and algal selected sequences appear to have diverged earlier than the land plant sequences, suggesting a more ancestral GHL17 origin (Gaudioso-Pedraza and Benítez-Alfonso, 2014). Other callose regulation proteins, such as the callose synthase (CalS) family, have been duplicated during the diversification of land plants (Drábková and Honys, 2017). Together, these findings indicate that plasmodesmata-localized proteins were already present in the land plant ancestor but that they played different roles.

As a consequence of plasmodesmata transport, plants can utilize the biogeneric properties of DPMs such as diffusion and lateral inhibition to specify cell identity and develop vascular tissues specialized in transporting nutrients over long distances. Without this capacity, plants would have not been able to generate the complex multicellular organisms that we know and that have become the major life form on earth. Despite the importance of plasmodesmata-mediated transport for plant development and diversification, little is known about their evolution. The reasons for this stem in part from the fact that plasmodesmata have structural characteristics that differ among different kinds of tissues as well as among the different plant lineages, and from the fact that the complete disruption of plant tissues is still challenging (Faulkner and Maule, 2011; Brunkard and Zambryski, 2017). However, some molecules that may be generically involved in the formation of plasmodesmata are now being postulated, as for example certain reticulons (Brunkard and Zambryski, 2017). It is likely that the advent of new methodologies that allow us to identify new plasmodesmata proteins will help elucidate the regulatory properties of plasmodesmata as well as the origins of these molecules and the genes encoding them in organisms that lack or that have less complex plasmodesmata.

MOLECULAR REGULATORY NETWORKS (MRNs): CO-OPTION, DRIFT AND PLANT EVOLUTIONARY TRANSITIONS

The notion of GRNs, recently referred to as molecular regulatory networks to include other types of molecules,

has allowed the fruitful exploration of the collective effect of genes and gene products in organismal development, although several other phenomena have been recently identified that call for a re-evaluation or update of current network modeling formalisms (e.g., the role of intrinsically disordered proteins in gene regulation, Niklas et al., 2015, 2018). Molecular regulatory networks integrate a set of nodes that can stand for genes, proteins, different types of RNA or other molecules, and a set of edges that correspond to the regulatory interactions among the elements represented by nodes. Multiple studies have aimed to study the dynamics of such networks, not only in plants, but also in animals, fungi and bacteria, mostly to test the idea that the steady states (attractors) of molecular regulatory networks correspond to specific cell types or cellular states (Kauffman, 1969; Thomas, 1991; Albert and Othmer, 2003; Alvarez-Buylla et al., 2007).

The picture emerging from theoretical and empirical studies is that molecular regulatory network steady states may indeed correspond to cell types or metabolic states, and that such different attractors can be present even in unicellular organisms that alternate between different phases or states in their life cycle (Quiñones-Valles et al., 2014; Mora Van Cauwelaert et al., 2015). However, the temporal coexistence of different cell types can only occur in multicellular organisms. Multiple studies have suggested that the molecular regulatory networks that underlie the specification of different cell-types in extant multicellular organisms may have been co-opted from multistable molecular regulatory networks, i.e., networks leading to more than one steady state, that were already present in their unicellular ancestors (Newman and Bhat, 2009; Mora Van Cauwelaert et al., 2015; Sebé-Pedrós et al., 2017). Indeed, mathematical and computational models have been used to perform proof-of-principle simulations that illustrate how single cells with multistable molecular regulatory networks can aggregate and couple via diverse communication mechanisms, giving rise to stereotypic and robust arrangements of cells with different identities (Furusawa and Kaneko, 2002; Mora Van Cauwelaert et al., 2015). This is a powerful idea, since this scenario requires no massive or abrupt genetic changes to explain one of the most major evolutionary transitions (Newman and Bhat, 2008, 2009; Niklas and Newman, 2013).

We have argued that some of the basic features of animal and plant body plans may have been generated by the cooption and differential spatiotemporal combination of DPMs (Newman and Bhat, 2008, 2009; Hernández-Hernández et al., 2012; Niklas and Newman, 2013; Niklas, 2014). However, DPMs are associated with molecules that are part of evolving regulatory networks such that DPM-related molecules and their regulatory interactions can change. As these networks evolve, DPMs may become canalized (*sensu* Waddington), this is, the patterns and shapes that were initially generated by generic physico-chemical processes mobilized by a few molecules can become somehow stabilized by the evolution of continuously more robust and intricate regulatory networks (Salazar-Ciudad et al., 2001). Molecular network evolution may also follow a trajectory characterized by

developmental system drift (DSD) (True and Haag, 2001), which suggests that genetic networks associated with phenotypes are both flexible and robust, and that differences between regulatory networks in related species arise by elimination or recruitment of new elements, erasing in this way traceable signals of common ancestry at the genetic level. DSD thus suggests that development of homologous traits in related species may not be mediated by homologous genetic factors (Müller and Newman, 1999; Rokas, 2006; Tsong et al., 2006; Kiontke et al., 2007; Nahmad and Lander, 2011; Sommer, 2012; Shbailat and Abouheif, 2013; Stolfi et al., 2014; Arias Del Angel et al., 2017). The mechanisms triggering divergence in the regulatory networks in DSD can involve both *cis*- and *trans*-regulatory changes, and the degree of change can vary from one system to another (Sommer, 2012; Stolfi et al., 2014).

We will now return to the DIF DPM example to illustrate some of these ideas.

All multicellular lineages with cellulosic cell walls appear to have evolved structures analogous to plasmodesmata. Indeed, as noted plasmodesmata evolved independently in different eukaryotic photosynthetic lineages and the molecules associated to their evolutionary origin are still unclear (Brunkard and Zambryski, 2017). However, some of the molecules that passively move through plasmodesmata and that are involved in the DIF DPM may have been co-opted from widely conserved molecular regulatory networks, some of which may predate plant multicellularity.

The case of auxin was briefly mentioned above. Indeed, currently available evidence shows that auxin biosynthesis was already present in the unicellular ancestors of multicellular eukaryotes (Beilby, 2016; Khasin et al., 2017; Ishizaki, 2017; Kato et al., 2017; among many other lines of evidence). However, this is not the case for the auxin transporters that have been identified and thoroughly studied in angiosperm model systems. It has thus been suggested that auxin initially moved only through plasmodesmata in a passive manner, contributing to multicellular organization through the formation of gradients and concentration patterns that could account for differential cellular behaviors and identities in vascular land plants, some algae, and bryophytes. If true, auxin transport seems to have been canalized and greatly potentialized by the evolution of complex molecular networks associated with its biosynthesis and transport. So much so, that in plants like *Arabidopsis* auxin local concentration is highly regulated and participates in diverse developmental processes and events under specific spatiotemporal conditions. Moreover, such tight regulation of auxin transport has enabled cellular and organ polarization, and has likely contributed to other evolutionary transitions, such as that to vascular plants. It also seems to be the case that extant networks associated with auxin biosynthesis and transport differ in particular elements and interactions, suggesting that the mechanisms of canalization have differed among plants or that some degree of developmental systems drift has occurred.

With regard to the DIF DPM, the role of molecular regulatory network cooption and further canalization is illustrated by the MYB-bHLH-WD40 protein complex. These plant proteins are

involved in complex networks that act in different plant organs and developmental stages, enabling the determination of diverse cell types, such as stomata, pavement cells, trichomes, and trichoblasts (Ramsay and Glover, 2005; Benítez et al., 2011; Torii, 2012; Horst et al., 2015; Breuninger et al., 2016). A central feature of this complex is that some of its components may move to neighboring cells through plasmodesmata, which gives rise to the coupling of otherwise intracellular networks and, concomitantly, the emergence of stereotypic cellular arrangements. Indeed, it is by the intercellular transport and mutual regulation of MYB, bHLH, and WD40 proteins that some of the well-known patterns of spaced-out stomata, trichomes, and aligned root hairs arise during plant development (Benítez et al., 2011; Torii, 2012; Horst et al., 2015). Interestingly, although the molecular regulatory networks in which these proteins take part seem to have come together in land plants, their key components appear to predate plant multicellularity (Ramsay and Glover, 2005). Consequently, some of the major events in the diversification of cellular types and functions have involved the co-option of ancient molecules, the presence of plasmodesmata, and the associated mobilization of certain DPMs, even if the MYB-bHLH-WD40 complex has drifted into regulatory systems that are currently species- or even organ-dependent.

FINAL REMARKS

Based on our review of the available evidence, we reach the following conclusions:

- Dynamical patterning modules, defined as sets of conserved gene products and molecular networks in conjunction with the physical morphogenetic and patterning processes they mobilize, have played ubiquitous and central roles in the evolution of multicellularity in the algae, land plants, fungi, as they have been shown to in metazoans (schematized in Figure 7).
- Four DPMs are critical in the context of multicellular evolution of plants and fungi: the DPM for the orientation of the future cell wall (FCW), the DPM for cell-to-cell adhesion (ADH), the DPM for polarity (POL), and the DPM for differentiation (DIF).
- Comparisons across the various fungal, algal, and land plant lineages indicate that these four DPMs have recruited different mechanisms and materials to achieve the same ends. For example, the materials used to achieve cell-to-cell adhesion differ dramatically among the various multicellular algal lineages and the land plants.
- These differences in the materials recruited by the various DPMs indicate that natural selection operates on morphological phenotypes and not on the components and mechanisms that produce them.

As exemplified in this study focusing on plant multicellularity, the DPM concept provides a valuable framework to further understand the processes behind multicellular development and evolution and can give rise to clear propositions that can in

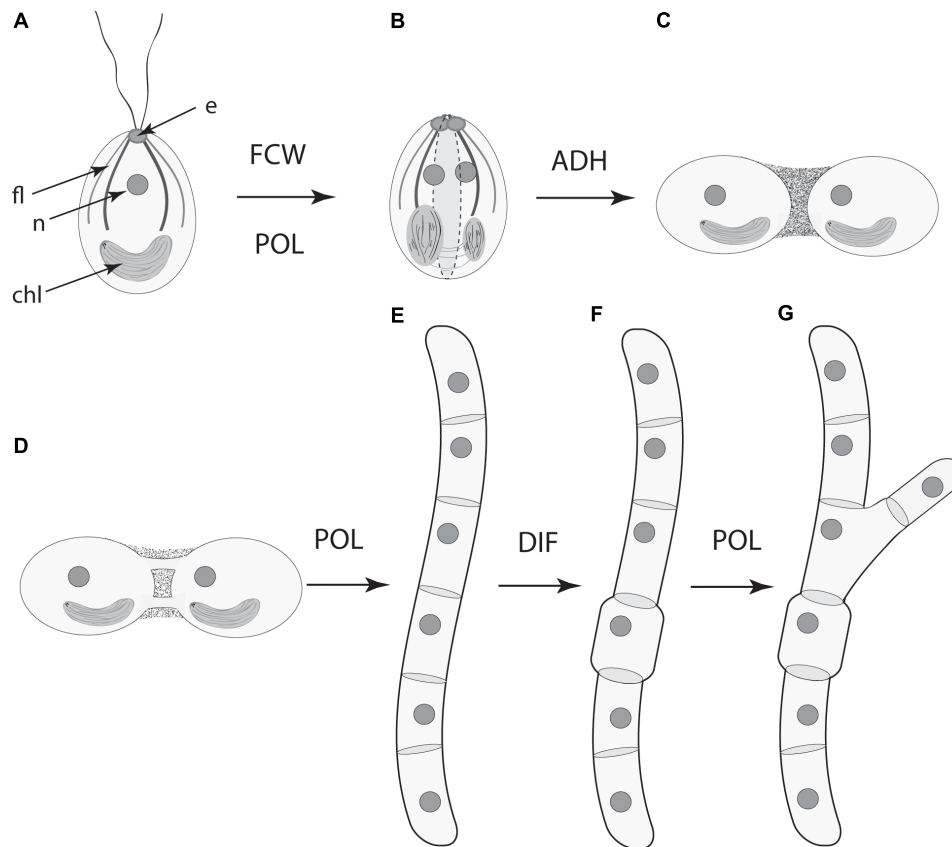


FIGURE 7 | Serial schematic of the ways four DPMs have contributed to the evolution of complex multicellularity in the land plants (see also, **Figure 5**). **(A)** The ancestral unicellular condition within the Streptophytes (the green algae and the land plants) with a flagellar apparatus and associated eyespot, nucleus, and parietal chloroplast. **(B)** The future cell wall (FCW) module operates along the longitudinal cell axis in conjunction with the polarity (POL) module to specify the plane of cell division (dashed lines). **(C)** The cell-to-cell adhesion (ADH) module establishes the colonial body plan (show here by two adjoining cells). **(D)** Symplastic continuity between adjoining cells is established by means of plasmodesmata. **(E)** The POL module shifts the frame of reference for the FCW module (now orthogonal to the cell longitudinal axis) to form an unbranched filament. **(F)** The differentiation (DIF) module is deployed for the evolution of specialized cells. **(G)** The POL module operates to now specify a second plane of cell division to produce a branched filament. The role of the POL module in establishing a third plane of cell division (and thus the development of parenchyma) is not shown here. e, eyespot; fl., flagellar apparatus; n, nucleus; chl, chloroplast.

turn be tested through comparative methods, mathematical and computational modeling, and experimental modification of parameters and biogeneric properties. However, the DPM concept has not been fully integrated into the “standard model” of contemporary evolutionary developmental biology. Typically, “mechanism” is considered at the level of genes and gene networks, while morphology is handled descriptively, with adaptationist narratives where they pertain, and appeals to pleiotropy and its consequences when they do not (Minelli, 2018). This perspective is unsatisfactory as an explanatory framework for biological form in light of the unquestioned role of physical mechanisms of morphogenesis across all categories of multicellular (Forgacs and Newman, 2005; Niklas and Newman, 2013; Newman and Niklas, 2018) and increased recognition of the conservation of early-evolved architectural motifs despite drift in molecular mechanisms (True and Haag, 2001).

Such homoplasy is even more pervasive in plant than in animal systems, where, as we have described here and elsewhere,

there have been multiple routes to multicellularity rather than the single, classical cadherin-based, one in the metazoans (Niklas and Newman, 2013; Newman, 2016b). Moreover, the ability of the cyanobacteria, the land plants, and the brown algae to form plasmodesmata-like intercellular structures involves significantly different GRNs, gene products, and developmental processes. Yet, the result in each case is the same, i.e., intercellular adhesion, communication, and polarity.

Whereas in animal systems GRNs and DPMs act relatively independently of each other, with the former mainly specifying cell type identity and latter patterns and arrangements of cells (Newman et al., 2009), the molecular regulatory networks of plants and fungi act in a more integrated fashion, comprising both GRN- and DPM-type functions of metazoans. This is partly because transcription factors move more freely between cells in the non-metazoans. Moreover, since the physics embodied in DPMs often leads to predictable morphological outcomes, these modules have served as “simplification forces” in evolution, acting as major instructive cues that channel development in both

animals and plants. In contrast to animal GRNs, however, the mixed-nature plant molecular regulatory networks have been “complexification forces” in plant and fungal evolution, offering additional opportunities to use/modulate/bridge DPMs to generate an enhanced spectrum of morphological complexity.

The behaviors of developing tissues as excitable biogenic materials (liquids and liquid crystals in the case of animals, deformable cellular solids in the case of plants), are inescapable, as are the preferred morphological motifs generated by characteristic DPMs of these materials, whatever their molecular genetic underpinnings may be. Understanding these inherent properties is essential to mechanistic explanations of development and its transformations during the evolution of multicellular organisms (Newman, 2017). A challenge for future research is to determine how these modules recruit and integrate the ancillary processes required to achieve the morphological variety seen across the broad phylogenetic spectrum of multicellular plants and fungi.

REFERENCES

- Alassimone, J., Naseer, S., and Geldner, N. (2010). A developmental framework for endodermal differentiation and polarity. *Proc. Natl. Acad. Sci. U.S.A.* 107, 5214–5219. doi: 10.1073/pnas.0910772107
- Albert, R., and Othmer, H. G. (2003). The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in *Drosophila melanogaster*. *J. Theor. Biol.* 223, 1–18. doi: 10.1016/S0022-5193(03)00035-3
- Alvarez-Buylla, E. R., Benítez, M., Dávila, E. B., Chaos, A., Espinosa-Soto, C., and Padilla-Longoria, P. (2007). Gene regulatory network models for plant development. *Curr. Opin. Plant Biol.* 10, 83–91. doi: 10.1016/j.pbi.2006.11.008
- Arias Del Angel, J. A., Escalante, A. E., Martínez-Castilla, L. P., and Benítez, M. (2017). An evo-devo perspective on multicellular development of Myxobacteria. *J. Exp. Zool. B Mol. Dev. Evol.* 328, 165–178. doi: 10.1002/jez.b.22727
- Beaumont, N. (2009). Modelling the transport of nutrients in early animals. *Evol. Biol.* 36, 256–266. doi: 10.1007/s11692-008-9047-2
- Beilby, M. J. (2016). Multi-scale characean experimental system: from electrophysiology of membrane transporters to cell-to-cell connectivity, cytoplasmic streaming and auxin metabolism. *Front. Plant Sci.* 7:1052. doi: 10.3389/fpls.2016.01052
- Benítez, M., and Hejatk, J. (2013). Dynamics of cell-fate determination and patterning in the vascular bundles of *Arabidopsis thaliana*. *PLoS One* 8:e63108. doi: 10.1371/journal.pone.0063108
- Benítez, M., Monk, N. A., and Alvarez-Buylla, E. R. (2011). Epidermal patterning in *Arabidopsis*: models make a difference. *J. Exp. Zool. B Mol. Dev. Evol.* 316, 241–253. doi: 10.1002/jez.b.21398
- Benítez-Alfonso, Y., Faulkner, C., Pendle, A., Miyashima, S., Helariutta, Y., and Maule, A. (2013). Symplastic intercellular connectivity regulates lateral root patterning. *Dev. Cell* 26, 136–147. doi: 10.1016/j.devcel.2013.06.010
- Bloemendal, S., and Kück, U. (2013). Cell-to-cell communication in plants, animals, and fungi: a comparative review. *Naturwissenschaften* 100, 3–19. doi: 10.1007/s00114-012-0988
- Bonner, J. T. (2000). *First Signals: The Evolution of Multicellular Development*. Princeton, NJ: Princeton University Press.
- Breuninger, H., Thamm, A., Streubel, S., Sakayama, H., Nishiyama, T., and Dolan, L. (2016). Diversification of a transcription factor family led to the evolution of antagonistically acting genetic regulators of root hair growth. *Curr. Biol.* 26, 1622–1628. doi: 10.1016/j.cub.2016.04.060
- Brunkard, J. O., and Zambryski, P. C. (2017). Plasmodesmata enable multicellularity: new insights into their evolution, biogenesis, and functions in development and immunity. *Curr. Opin. Plant Biol.* 35, 76–83. doi: 10.1016/j.pbi.2016.11.007
- Cannon, M. C., Terneus, K., Hall, Q., Tan, L., Wang, Y., Wegenhart, B. L., et al. (2008). Self-assembly of the plant cell wall requires an extension scaffold. *Proc. Natl. Acad. Sci. U.S.A.* 105, 2226–2231. doi: 10.1073/pnas.0711980105
- Carlsbecker, A., Lee, J. Y., Roberts, C. J., Dettmer, J., Lehesranta, S., Zhou, J., et al. (2010). Cell signalling by microRNA165/6 directs gene dose-dependent root cell fate. *Nature* 465, 316–321. doi: 10.1038/nature08977
- Christensen, N. M., Faulkner, C., and Oparka, K. (2009). Evidence for unidirectional flow through plasmodesmata. *Plant Physiol.* 150, 96–104. doi: 10.1104/pp.109.137083
- Cook, M., Graham, L., Botha, C., and Lavin, C. (1997). Comparative ultrastructure of plasmodesmata of Chara and selected bryophytes: toward an elucidation of the evolutionary origin of plant plasmodesmata. *Am. J. Bot.* 84:1169. doi: 10.2307/2446040
- Cook, M., Graham, L., Van Bel, A. J. E., and van Kesteren, W. J. P. (1999). “Evolution of plasmodesmata,” in *Plasmodesmata: Structure, Function, Role in Cell Communication*, eds A. J. E. van Bel and W. J. P. Kesteren (Berlin: Springer).
- Cui, H., Levesque, M. P., Vernoux, T., Jung, J. W., Paquette, A. J., Gallagher, K. L., et al. (2007). An evolutionarily conserved mechanism delimiting SHR movement defines a single layer of endodermis in plants. *Science* 316, 421–425. doi: 10.1126/science.1139531
- De Storme, N., and Geelen, D. (2014). Callose homeostasis at plasmodesmata: molecular regulators and developmental relevance. *Front. Plant Sci.* 5:138. doi: 10.3389/fpls.2014.00138
- Dettmer, J., and Friml, J. (2011). Cell polarity in plants: when two do the same, it is not the same. *Curr. Opin. Cell Biol.* 23, 686–696. doi: 10.1016/j.cob.2011.09.006
- Drábková, L. Z., and Honys, D. (2017). Evolutionary history of callose synthases in terrestrial plants with emphasis on proteins involved in male gametophyte development. *PLoS One* 12:e0187331. doi: 10.1371/journal.pone.0187331
- Faulkner, C., and Maule, A. (2011). Opportunities and successes in the search for plasmodesmal proteins. *Protoplasma* 248, 27–38. doi: 10.1007/s00709-010-0213-x
- Ferris, P. J., Woessner, J. P., Waffenschmidt, S., Kilz, S., Drees, J., and Goodenough, U. W. (2001). Glycosylated polyproline II rods with kinks as a structural motif in plant hydroxyproline-rich glycoproteins. *Biochemistry* 40, 2978–2987. doi: 10.1021/bi0023605
- Fleury, V. (1999). Un possible lien entre la croissance dendritique en physique et la morphogenèse des plantes. *C. R. Acad. Sci. Paris Ser. III Sci. Vie* 322, 725–734. doi: 10.1016/S0764-4469(00)80030-X
- Floyd, G. L., Stewart, K. D., and Mattox, K. R. (1971). Cytokinesis and plasmodesmata in *Ulothrix*. *J. Phycol.* 7, 306–309.
- Folse, H. J. Jr., and Roughgarden, J. (2010). What is an individual organism? A multilevel selection perspective. *Q. Rev. Biol.* 85, 447–472. doi: 10.1086/656905

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

Funding from the College of Agriculture and Life Sciences (to KN) and CONACyt (to VH-H) is gratefully acknowledged.

ACKNOWLEDGMENTS

The authors thank Drs. Verónica S. Di Stilio (University of Washington), Annette Becker (Justus-Liebig-Universität Gießen), and Natalia Pabón-Mora (Universidad de Antioquia Medellín) for inviting this contribution and reviewers for their perceptive and constructive comments.

- Forgacs, G., and Newman, S. A. (2005). *Biological Physics of the Developing Embryo*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511755576
- Fraser, T. W., and Gunning, B. E. S. (1969). The ultrastructure of plasmodesmata in the filamentous green alga, *Bulbochaete hiloensis* (Nordst.) tiffany. *Planta* 88, 244–254. doi: 10.1007/BF00385067
- Furusawa, C., and Kaneko, K. (2002). Origin of multicellular organisms as an inevitable consequence of dynamical systems. *Anat. Rec.* 268, 327–342. doi: 10.1002/ar.10164
- Galletti, R., Verger, S., Hamant, O., and Ingram, G. C. (2016). Developing a 'thick skin': a paradoxical role for mechanical tension in maintaining epidermal integrity? *Development* 143, 3249–3258. doi: 10.1242/dev.132837
- Gardiner, J. (2013). The evolution and diversification of plant microtubule-associated proteins. *Plant J.* 75, 219–229. doi: 10.1111/tpj.12189
- Garrett, J. J., Meets, M. J., Blackshaw, M. T., Blackshaw, L. C., Hou, H., Styranke, D. M., et al. (2012). A novel, semi-dominant allele of MONOPTEROS provide insight into leaf initiation and vein pattern formation. *Planta* 236, 297–312. doi: 10.1007/s00425-012-1607-0
- Gaudioso-Pedraza, R., and Benítez-Alfonso, Y. (2014). A phylogenetic approach to study the origin and evolution of plasmodesmata-localized glycosyl hydrolases family 17. *Front. Plant Sci.* 5:212. doi: 10.3389/fpls.2014.00212
- Geldner, N. (2009). Cell polarity in plants—A PARspective on PINs. *Curr. Opin. Plant Biol.* 12, 42–48. doi: 10.1016/j.pbi.2008.09.009
- Graham, L. E., Graham, J. M., and Wilcox, L. W. (2009). *Algae*, 2nd Edn. San Francisco, CA: Benjamin Cummings.
- Green, K. J., Viamontes, G. I., and Kirk, D. L. (1981). Mechanism of formation, ultrastructure and function of the cytoplasmic bridge system during morphogenesis in *Volvox*. *J. Cell Biol.* 91, 756–769. doi: 10.1083/jcb.91.3.756
- Green, S., and Batterman, R. (2017). Biology meets physics: reductionism and multi-scale modeling of morphogenesis. *Stud. Hist. Philos. Biol. Biomed. Sci.* 61, 20–34. doi: 10.1016/j.shpsc.2016.12.003
- Grosberg, R. K., and Strathmann, R. (2007). The evolution of multicellularity: a minor major transition? *Annu. Rev. Ecol. Syst.* 38, 621–654. doi: 10.1146/annurev.ecolsys.36.102403.114735
- Guenounne-Gelbart, D., Elbaum, M., Sagi, G., Levy, A., and Epel, B. L. (2008). Tobacco mosaic virus (TMV) replicase and movement protein function synergistically in facilitating TMV spread by lateral diffusion in the plasmodesmal desmotubule of *Nicotiana benthamiana*. *Mol. Plant Microbe Interact.* 21, 335–345. doi: 10.1094/MPMI-21-3-0335
- Guseman, J. M., Lee, J. S., Bogenschutz, N. L., Peterson, K. M., Virata, R. E., Xie, B., et al. (2010). Dysregulation of cell-to-cell connectivity and stomatal patterning by loss-of-function mutation in *Arabidopsis* CHORUS (GLUCAN SYNTHASE-LIKE 8). *Development* 137, 1731–1741. doi: 10.1242/dev.049197
- Hall, J. D., McCourt, R. M., and Delwiche, C. F. (2008). Patterns of cell division in the filamentous Desmidiaceae, close green algal relatives of land plants. *Am. J. Bot.* 95, 643–654. doi: 10.3732/ajb.2007210
- Han, X., Hyun, T. K., Zhang, M., Kumar, R., Koh, E. J., Kang, B. H., et al. (2014). Auxin-callose-mediated plasmodesmal gating is essential for tropic auxin gradient formation and signaling. *Dev. Cell* 28, 132–146. doi: 10.1016/j.devcel.2013.12.008
- Hernández-Hernández, V., Niklas, K. J., Newman, S. A., and Benítez, M. (2012). Dynamical patterning modules in plant development and evolution. *Int. J. Dev. Biol.* 56, 661–674. doi: 10.1387/ijdb.120027mb
- Herron, M. D., and Michod, R. E. (2008). Evolution of complexity in the volvocine algae: transitions in individuality through Darwin's eye. *Evolution* 62, 436–451. doi: 10.1111/j.1558-5646.2007.00304.x
- Hoops, H. J., Nishii, I., and Kirk, D. L. (2000). "Cytoplasmic bridges in *Volvox* and its relatives," in *Madame Curie Bioscience Database [Internet]* (Austin, TX: Landes Bioscience). Available at: <http://www.ncbi.nlm.nih.gov/books/NBK6424/>
- Horst, R. J., Fujita, H., Lee, J. S., Rychel, A. L., Garrick, J. M., Kawaguchi, M., et al. (2015). Molecular framework of a regulatory circuit initiating two-dimensional spatial patterning of stomatal lineage. *PLoS Genet.* 11:e1005374. doi: 10.1371/journal.pgen.1005374
- Ishizaki, K. (2017). Evolution of land plants: insights from molecular studies on basal lineages. *Biosci. Biotechnol. Biochem.* 81, 73–80. doi: 10.1080/09168451.2016.1224641
- Kato, H., Nishihama, R., Weijers, D., and Kohchi, T. (2017). Evolution of nuclear auxin signaling: lessons from genetic studies with basal land plants. *J. Exp. Bot.* 69, 291–301. doi: 10.1093/jxb/erx267
- Kauffman, S. A. (1969). Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theor. Biol.* 22, 437–467. doi: 10.1016/0022-5193(69)90015-0
- Khasin, M., Cahoon, R. R., Nickerson, K. W., and Riekhof, W. R. (2017). Molecular machinery of auxin synthesis, secretion, and perception in the unicellular chlorophyte alga *Chlorella sorokiniana* UTEX 1230. *bioRxiv* doi: 10.1101/172833 [Preprint].
- Kiontke, K., Barriere, A., Kolotuev, I., Podbilewicz, B., Sommer, R., Fitch, D. H., et al. (2007). Trends, stasis, and drift in the evolution of nematode vulva development. *Curr. Biol.* 17, 1925–1937. doi: 10.1016/j.cub.2007.10.061
- Kirk, D. L. (2005). A twelve-step program for evolving multicellularity and a division of labor. *Bioessays* 27, 299–310. doi: 10.1002/bies.20197
- Kitagawa, M., and Fujita, T. (2013). Quantitative imaging of directional transport through plasmodesmata in moss protonemata via single-cell photoconversion of Dendra2. *J. Plant Res.* 126, 577–585. doi: 10.1007/s10265-013-0547-5
- Knoll, A. (2011). The multiple origins of complex multicellularity. *Annu. Rev. Earth Planet. Sci.* 39, 217–239. doi: 10.1146/annurev.earth.031208.100209
- Kutschera, U., and Niklas, K. J. (2005). Endosymbiosis, cell evolution, and speciation. *Theory Biosci.* 124, 1–24. doi: 10.1016/j.thbio.2005.04.001
- Kutschera, U., and Niklas, K. J. (2007). The epidermal-growth-control theory of stem elongation: an old and new perspective. *Plant Physiol.* 164, 1395–1409. doi: 10.1016/j.jplph.2007.08.002
- Kutschera, U., and Niklas, K. J. (2008). Macroevolution via secondary endosymbiosis: a Neo-Goldschmidtian view of unicellular hopeful monsters and Darwin's primordial intermediate form. *Theory Biosci.* 127, 277–289. doi: 10.1007/s12064-008-0046-8
- Kwiatkowska, M. (1999). "Plasmodesmal coupling and cell differentiation in algae," in *Plasmodesmata: Structure, Function, Role in Cell Communication*, eds A. J. E. Van Bel and W. J. P. van Kesteren (Berlin: Springer).
- Lamport, D. T. A., Kieliszewski, M. J., Chen, Y., and Cannon, M. C. (2011). Role of the extension superfamily in primary cell wall architecture. *Plant Physiol.* 156, 11–19. doi: 10.1104/pp.110.169011
- Levine, H., and Ben-Jacob, E. (2004). Physical schemata underlying biological pattern formation—examples, issues and strategies. *Phys. Biol.* 1, P14–P22. doi: 10.1088/1478-3967/1/2/P01
- Lucas, W. I., Ding, B., and Van der Schoot, C. (1993). Plasmodesmata and the supracellular nature of plants. *New Phytol.* 125, 435–476. doi: 10.1111/j.1469-8137.1993.tb03897.x
- Lund, E. J. (1947). *Bioelectric Fields and Growth*. Austin, TX: University of Texas Press.
- Majda, M., Grones, P., Sinton, I.-M., Vain, T., Milani, P., Krupinski, P., et al. (2017). Mechanochemical polarization of contiguous cell walls shapes plant pavement cells. *Dev. Cell* 43, 290–304. doi: 10.1016/j.devcel.2017.10.017
- Marchant, H. J. (1976). "Plasmodesmata in algae and fungi," in *Intercellular Communication in Plants: Studies on Plasmodesmata*, eds B. E. S. Gunning and A. W. Robards (New York, NY: Springer), 59–78.
- McGhee, G. R. (1999). *Theoretical Morphology: The Concept and Its Applications*. New York, NY: Columbia University Press.
- Menzel, D. (1996). The role of the cytoskeleton in polarity and morphogenesis of algal cells. *Curr. Opin. Cell Biol.* 1, 38–42. doi: 10.1016/S0955-0674(96)80046-9
- Meyerowitz, E. M. (2002). Plants compared to animals: the broadest comparative study of development. *Science* 295, 1482–1485. doi: 10.1126/science.1066609
- Minelli, A. (2018). *Plant Evolutionary Developmental Biology: The Evolvability of the Phenotype*. Cambridge: Cambridge University Press. doi: 10.1017/9781139542364
- Mora Van Cauwelaert, E., Del Angel, A., Antonio, J., Benítez, M., and Azpeitia, E. M. (2015). Development of cell differentiation in the transition to multicellularity: a dynamical modeling approach. *Front. Microbiol.* 6:603. doi: 10.3389/fmicb.2015.00603
- Müller, G. B., and Newman, S. A. (1999). "Generation, integration, autonomy: three steps in the evolution in homology," in *Homology*, Vol. 222, eds G. K. Bock and G. Cardew (Chichester: Wiley), 65–73.
- Nahmad, M., and Lander, A. D. (2011). Spatiotemporal mechanisms of morphogen gradient interpretation. *Curr. Opin. Genet. Dev.* 21, 726–731. doi: 10.1016/j.gde.2011.10.002

- Newman, S. A. (2011). The developmental specificity of physical mechanisms. *Ludus Vitalis* 19, 343–351.
- Newman, S. A. (2012). Physico-genetic determinants in the evolution of development. *Science* 338, 217–219. doi: 10.1126/science.1222003
- Newman, S. A. (2016a). 'Biogenic' developmental processes: drivers of major transitions in animal evolution. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 371:20150443. doi: 10.1098/rstb.2015.0443
- Newman, S. A. (2016b). "Multicellularity, the emergence of animal body plans, and the stabilizing role of the egg," in *Multicellularity: Origins and Evolution*, eds K. J. Niklas and S. A. Newman (Cambridge, MA: MIT Press), 225–246.
- Newman, S. A. (2017). "Inherency," in *Evolutionary Developmental Biology*, eds L. Nuño de la Rosa and G. B. Müller (Basel: Springer). doi: 10.1007/978-3-319-33038-9_78-1
- Newman, S. A., and Bhat, R. (2008). Dynamical patterning modules: physico-genetic determinants of morphological development and evolution. *Phys. Biol.* 5:15008. doi: 10.1088/1478-3975/5/1/015008
- Newman, S. A., and Bhat, R. (2009). Dynamical patterning modules: a "pattern language" for development and evolution of multicellular form. *Int. J. Dev. Biol.* 53, 693–705. doi: 10.1387/ijdb.072481sn
- Newman, S. A., Bhat, R., and Mezentsseva, N. V. (2009). Cell state switching factors and dynamical patterning modules: complementary mediators of plasticity in development and evolution. *J. Biosci.* 34, 553–572. doi: 10.1007/s12038-009-0074-7
- Newman, S. A., and Niklas, K. J. (2018). "Dynamical patterning modules link genotypes to morphological phenotypes in multicellular evolution," in *Cells in Evolutionary Biology*, eds B. K. Hall and S. A. Moody (Boca Raton, FL: CRC Press), 235–266.
- Niklas, K. J. (1989). Mechanical behavior of plant tissues as inferred from the theory of pressurized cellular solids. *Am. J. Bot.* 76, 929–937. doi: 10.1002/j.1537-2197.1989.tb15071.x
- Niklas, K. J. (1992). *Plant Biomechanics*. Chicago, IL: University of Chicago Press.
- Niklas, K. J. (1994). *Plant Allometry*. Chicago, IL: University of Chicago Press.
- Niklas, K. J. (2000). The evolution of plant body plans — A biomechanical perspective. *Ann. Bot.* 85, 411–438. doi: 10.1006/anbo.1999.1100
- Niklas, K. J. (2014). The evolutionary-developmental origins of multicellularity. *Am. J. Bot.* 101, 6–25. doi: 10.3732/ajb.1300314
- Niklas, K. J. (2017). Size-dependent variation in plant form. *Curr. Biol.* 27, R853–R909. doi: 10.1016/j.cub.2017.02.007
- Niklas, K. J., Bondos, S. E., Dunker, A. K., and Newman, S. A. (2015). Rethinking gene regulatory network theory in light of alternative splicing, intrinsically disordered protein domains, and post-translational modifications. *Front. Cell Dev. Biol.* 3:8. doi: 10.3389/fcell.2015.00008
- Niklas, K. J., Cobb, E. D., and Crawford, D. R. (2013). The evo-devo of multicellular cells, tissues and organisms, and an alternative route to multicellularity. *Evol. Dev.* 15, 466–474. doi: 10.1111/ede.12055
- Niklas, K. J., Cobb, E. D., and Dunker, A. K. (2014). The number of cell types, information content, and the evolution of multicellularity. *Acta Soc. Bot. Pol.* 83, 337–347. doi: 10.5586/asbp.2014.034
- Niklas, K. J., Dunker, A. K., and Yruea, I. (2018). The evolutionary origins of cell type diversification and the role of intrinsically disordered proteins. *J. Exp. Bot.* 69, 1427–1446. doi: 10.1093/jxb/erx493
- Niklas, K. J., and Kaplan, D. R. (1991). "Biomechanics and the adaptive significance of multicellularity in plants," in *Proceedings of the Fourth International Congress of Systematics and Evolutionary Biology: The Unity of Evolutionary Biology*, ed. E. C. Dudley (Portland, OR: Dioscorides Press), 489–502.
- Niklas, K. J., and Newman, S. A. (2013). The origins of multicellular organisms. *Evol. Dev.* 15, 41–52. doi: 10.1111/ede.12013
- Niklas, K. J., and Spatz, H.-C. (2012). *Plant Physics*. London: University of Chicago Press. doi: 10.7208/chicago/9780226586342.001.0001
- Peaucelle, A., Wightman, R., and Höfte, H. (2015). The control of growth symmetry breaking in the *Arabidopsis* hypocotyl. *Curr. Biol.* 25, 1746–1752. doi: 10.1016/j.cub.2015.05.022
- Pillitteri, L. J., and Dong, J. (2013). Stomatal development in *Arabidopsis*. *Arabidopsis Book* 11:e0162. doi: 10.1199/tab.0162
- Prochiantz, A. (2011). Homeoprotein intercellular transfer, the hidden face of cell—Penetrating peptides. *Methods Mol. Biol.* 683, 249–257. doi: 10.1007/978-1-60761-919-2_18
- Quiñones-Valles, C., Sánchez-Osorio, I., and Martínez-Antonio, A. (2014). Dynamical modeling of the cell cycle and cell fate emergence in *Caulobacter crescentus*. *PLoS One* 9:e111116. doi: 10.1371/journal.pone.0111116
- Ramsay, N. A., and Glover, B. J. (2005). MYB–bHLH–WD40 protein complex and the evolution of cellular diversity. *Trends Plant Sci.* 10, 63–70. doi: 10.1016/j.tplants.2004.12.011
- Ranjan, A., Townsley, B. T., Ichihashi, Y., Sinha, N. R., and Chitwood, R. H. (2015). An intracellular transcriptomic atlas of the giant coenocyte *Caulerpa taxifolia*. *PLoS Genet.* 11:e1004900. doi: 10.1371/journal.pgen.1004900
- Raven, J. A. (1997). Miniview: multiple origins of plasmodesmata. *Eur. J. Phycol.* 32, 95–101. doi: 10.1080/09670269710001737009
- Rokas, A. (2006). Evolution: different paths to the same end. *Nature* 443, 401–402. doi: 10.1038/443401a
- Ruan, Y. L., Xu, S. M., White, R., and Furbank, R. T. (2004). Genotypic and developmental evidence for the role of plasmodesmata regulation in cotton fiber elongation mediated by callose turnover. *Plant Physiol.* 136, 4104–4113. doi: 10.1104/pp.104.051540
- Sager, R., and Lee, J. Y. (2014). Plasmodesmata in integrated cell signalling: insights from development and environmental signals and stresses. *J. Exp. Bot.* 65, 6337–6358. doi: 10.1093/jxb/eru365
- Salazar-Ciudad, I., Newman, S. A., and Solé, R. (2001). Phenotypic and dynamical transitions in model genetic networks. I. Emergence of patterns and genotype-phenotype relationships. *Evol. Dev.* 3, 84–94. doi: 10.1046/j.1525-142x.2001.003002084.x
- Schaefer, E., Belcram, K., Uyttewaald, M., Duroc, Y., Goussot, M., Legland, D., et al. (2017). The preprophase band of microtubules controls the robustness of division orientation in plants. *Proc. Natl. Acad. Sci. U.S.A.* 356, 186–189. doi: 10.1126/science.aal3016
- Scherp, P., Grotha, R., and Kutschera, U. (2001). Occurrence and phylogenetic significance of cytokinesis-related callose in green algae, bryophytes, ferns and seed plants. *Plant Cell Rep.* 20, 143–149. doi: 10.1007/s002990000301
- Schuette, S., Wood, A. J., Geisler, M., Geisler-Lee, J., Ligrone, R., and Renzaglia, K. S. (2009). Novel localization of callose in the spores of *Physcomitrella patens* and phylogenomics of the callose synthase gene family. *Ann. Bot.* 103, 749–756. doi: 10.1093/aob/mcn268
- Sebé-Pedrós, A., Degnan, B. M., and Ruiz-Trillo, I. (2017). The origin of Metazoa: a unicellular perspective. *Nat. Rev. Genet.* 18, 498–512. doi: 10.1038/nrg.2017.21
- Seiler, S., and Justa-Schuch, D. (2010). Conserved components, but distinct mechanisms for the placement and assembly of the cell division machinery in unicellular and filamentous ascomycetes. *Mol. Microbiol.* 78, 1058–1076. doi: 10.1111/j.1365-2958.2010.07392.x
- Shbailat, S. J., and Abouheif, E. (2013). The wing-patterning network in the wingless castes of Myrmicine and Formicine ant species is a mix of evolutionarily labile and non-labile genes. *J. Exp. Zool. B Mol. Dev. Evol.* 320, 74–83. doi: 10.1002/jez.b.22482
- Shin, K., Fogg, V. C., and Margolis, B. (2006). Tight junctions and cell polarity. *Annu. Rev. Cell Dev. Biol.* 22, 207–235. doi: 10.1146/annurev.cellbio.22.010305.104219
- Sideman, E. J., and Scheirer, D. C. (1977). Some fine structural observations on developing and mature sieve elements in the brown alga *Laminaria saccharina*. *Am. J. Bot.* 64, 649–657. doi: 10.1002/j.1537-2197.1977.tb11905.x
- Sinha, S., and Sridhar, S. (2015). *Patterns in Excitable Media: Genesis, Dynamics, and Control*. Boca Raton, FL: CRC Press.
- Sommer, R. J. (2012). Evolution of regulatory networks: nematode vulva induction as an example of developmental systems drift. *Adv. Exp. Med. Biol.* 751, 79–91. doi: 10.1007/978-1-4614-3567-9_4
- Stanley, M. S., Perry, R. M., and Callow, J. A. (2005). Analysis of expressed sequence tags (ESTs) from the green alga *Ulva linza* (Chlorophyta). *J. Phycol.* 41, 1219–1226. doi: 10.1111/j.1529-8817.2005.00138.x
- Stewart, K. D., Mattox, K. R., and Floyd, G. L. (1973). Mitosis, cytokinesis, the distribution of plasmodesmata, and other cytological characteristics in the Ulotrichales, Ulvales, and Chaetophorales: phylogenetic and taxonomic considerations. *J. Phycol.* 9, 128–141. doi: 10.1111/j.1529-8817.1973.tb04068.x
- Stolfi, A., Lowe, E. K., Racioppi, C., Ristatore, F., Brown, C. T., Swalla, B. J., et al. (2014). Divergent mechanisms regulate conserved cardiopharyngeal development and gene expression in distantly related ascidians. *eLife* 3:e03728. doi: 10.7554/eLife.03728

- Sumper, M., and Hallmann, A. (1998). Biochemistry of the extracellular matrix of *Volvox*. *Int. Rev. Cytol.* 180, 51–85. doi: 10.1016/S0074-7696(08)61770-2
- Swift, J., Ivanovska, I. L., Buxboim, A., Harada, T., Dingal, P. C., Pinter, J., et al. (2013). Nuclear lamin-A scales with tissue stiffness and enhances matrix-directed differentiation. *Science* 341:1240104. doi: 10.1126/science.1240104
- Thomas, R. (1991). Regulatory networks seen as asynchronous automata: a logical description. *J. Theor. Biol.* 153, 1–23. doi: 10.1016/S0022-5193(05)80350-9
- Torii, K. U. (2012). Two-dimensional spatial patterning in developmental systems. *Trends Cell Biol.* 22, 438–446. doi: 10.1016/j.tcb.2012.06.002
- True, J. R., and Haag, E. S. (2001). Developmental system drift and flexibility in evolutionary trajectories. *Evol. Dev.* 3, 109–119. doi: 10.1046/j.1525-142x.2001.003002109.x
- Tsong, A. E., Tuch, B. B., Li, H., and Johnson, A. D. (2006). Evolution of alternative transcriptional circuits with identical logic. *Nature* 443, 415–420. doi: 10.1038/nature05099
- Tylewicz, S., Petterle, A., Marttila, S., Miskolczi, P., Azeez, A., Singh, R. K., et al. (2018). Photoperiodic control of seasonal growth is mediated by ABA acting on cell-cell communication. *Science* 360, 212–214. doi: 10.1126/science.aan8576
- Urbanus, S. L., Martinelli, A. P., Dinh, Q. D. P., Aizza, L. C. B., Dornelas, M. C., Angenent, G. C., et al. (2010). Intercellular transport of epidermis-expressed MADS domain transcription factors and their effect on plant morphology and floral transition. *Plant J.* 63, 60–72. doi: 10.1111/j.1365-313X.2010.04221.x
- Vatén, A., Dettmer, J., Wu, S., Stierhof, Y. D., Miyashima, S., Yadav, S. R., et al. (2011). Callose biosynthesis regulates symplastic trafficking during root development. *Dev. Cell* 21, 1144–1155. doi: 10.1016/j.devcel.2011.10.006
- Vergara-Silva, F., Martínez-Castilla, L., and Alvarez-Buylla, E. R. (2000). MADS-box genes: development and evolution of plant body plans. *J. Phycol.* 36, 803–812. doi: 10.1046/j.1529-8817.2000.03654.x
- Wayne, R. (2009). *Plant Cell Biology*. Amsterdam: Elsevier.
- Xoconostle-Cazares, B., Yu, X., Ruiz-Medrano, R., Wang, H. L., Monzer, J., Yoo, B.-C., et al. (1999). Plant paralog to viral movement protein that potentiates transport of mRNA into the phloem. *Science* 283, 94–98. doi: 10.1126/science.283.5398.94
- Yruela, I., Oldfield, C. J., Niklas, K. J., and Dunker, A. K. (2017). Evidence for a strong correlation between transcription factor protein disorder and organismic complexity. *Genome Biol. Evol.* 9, 1248–1265. doi: 10.1093/gbe/evx073
- Zažímalová, E., Murphy, A. S., Yang, H., Hoyerová, K., and Hošek, P. (2010). Auxin transporters—why so many? *Cold Spring Harb. Perspect. Biol.* 2:a001552. doi: 10.1101/cshperspect.a001552
- Zhang, X., Facette, M., Humphries, J. A., Shen, Z., Park, Y., Sutimantanapi, D., et al. (2012). Identification of PAN2 by quantitative proteomics as a leucine-rich repeat-receptor-like kinase acting upstream of PAN1 to polarize cell division in maize. *Plant Cell* 24, 4577–4589. doi: 10.1105/tpc.112.104125

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 © 2018 Benítez, Hernández-Hernández, Newman and Niklas. 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 Protein Ductility in Duplicated Genes of Plants

Inmaculada Yruela^{1,2*}, Bruno Contreras-Moreira^{1,2,3}, A. Keith Dunker⁴ and Karl J. Niklas⁵

¹ Estación Experimental de Aula Dei, Consejo Superior de Investigaciones Científicas, Zaragoza, Spain, ² Group of Biochemistry, Biophysics and Computational Biology, Joint Unit to CSIC, Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, Zaragoza, Spain, ³ Fundación Agencia Aragonesa para la Investigación y el Desarrollo (ARAID), Zaragoza, Spain, ⁴ Department of Biochemistry and Molecular Biology, Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, IN, United States, ⁵ Plant Biology Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, United States

OPEN ACCESS

Edited by:

Verónica S. Di Stilio,
University of Washington,
United States

Reviewed by:

Uener Kolukisaoglu,
Universität Tübingen, Germany
Lena Hileman,
The University of Kansas,
United States

*Correspondence:

Inmaculada Yruela
i.yruela@csic.es

Specialty section:

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

Received: 04 May 2018

Accepted: 30 July 2018

Published: 20 August 2018

Citation:

Yruela I, Contreras-Moreira B,
Dunker AK and Niklas KJ (2018)
Evolution of Protein Ductility
in Duplicated Genes of Plants.
Front. Plant Sci. 9:1216.
doi: 10.3389/fpls.2018.01216

Previous work has shown that ductile/intrinsically disordered proteins (IDPs) and residues (IDRs) are found in all unicellular and multicellular organisms, wherein they are essential for basic cellular functions and complement the function of rigid proteins. In addition, computational studies of diverse phylogenetic lineages have revealed: (1) that protein ductility increases in concert with organismic complexity, and (2) that distributions of IDPs and IDRs along the chromosomes of plant species are non-random and correlate with variations in the rates of the genetic recombination and chromosomal rearrangement. Here, we show that approximately 50% of aligned residues in paralogs across a spectrum of algae, bryophytes, monocots, and eudicots are IDRs and that a high proportion (ca. 60%) are in disordered segments greater than 30 residues. When three types of IDRs are distinguished (i.e., identical, similar and variable IDRs) we find that species with large numbers of chromosome and endoduplicated genes exhibit paralogous sequences with a higher frequency of identical IDRs, whereas species with small chromosomes numbers exhibit paralogous sequences with a higher frequency of similar and variable IDRs. These results are interpreted to indicate that genome duplication events influence the distribution of IDRs along protein sequences and likely favor the presence of identical IDRs (compared to similar IDRs or variable IDRs). We discuss the evolutionary implications of gene duplication events in the context of ductile/disordered residues and segments, their conservation, and their effects on functionality.

Keywords: IDPs, polyploidy, protein ductility, protein disorder, paralogs, genome duplication, plants

INTRODUCTION

There is wide consensus that spontaneous whole genome duplications (WGD, autopolyploidy) and interspecific hybridization (allopolyploidy), followed by post-polyploid diploidization (PPD) have contributed significantly to the evolution of the land plants and to the angiosperms in particular (Wendel, 2000; Ramsey and Schemske, 2002; Soltis and Soltis, 2009; Jackson and Chen, 2010; Mandáková and Lysak, 2018; Ren et al., 2018). In general, new species emerging from either type of polyploidy tend to exhibit improved growth vigor and adaptive resilience to adverse environments thereby conferring significant evolutionary advantages (Song and Chen, 2015). Although the

reasons remain unclear, plant genomes tend to have larger numbers of duplicated genes compared with the genomes of non-photosynthetic eukaryotes, although recent reports suggest that WGD events have also been frequent in insects (Li et al., 2018). Among the angiosperms, there is evidence that major clade-wide WGD events have occurred multiple times over the past 200 Mya (Lyons et al., 2008; Renny-Byfield and Wendel, 2014; del Pozo and Ramírez-Parra, 2015; Landis et al., 2018; Ren et al., 2018) in contrast to duplication events within major vertebrate lineages (Panopoulou et al., 2003; Dehal and Boore, 2005). In addition, a whole genome triplication events (triploidization, WGT, or hexaploidization) occurred in the ancestor of the core eudicots (approximately 125 Mya) and another more recent event (between 23 and 47 Mya) occurred in *Brassica* species, which affected the evolution of many agriculturally important species (Zheng et al., 2013; Kagale et al., 2014; Parkin et al., 2014). Thus, ancient and recent autopolyploidy have profoundly influenced the evolution of the flowering plants and have contributed to improved important agronomic traits, such as grain quality, fruit shape, and flowering time (Kellogg and Bennetzen, 2004; Dubcovsky and Dvorak, 2007; Leitch and Leitch, 2008; Jackson and Chen, 2010; Panchy et al., 2016). Given the fact that PPD events are recurrent over the course of angiosperm evolution, many extant diploid genomes contain a record of ancient WGD events that can be inferred by the analysis of duplicated genes with conserved co-linearity across genomic segments (Proost et al., 2012; Ren et al., 2018).

Despite their overarching importance, the consequences of polyploidy remain poorly understood. Studies have documented rapid and dynamic changes in genomic structure and gene expression in plant polyploids, which reflect genomic and functional plasticity of duplicated genes (Dubcovsky and Dvorak, 2007; Leitch and Leitch, 2008; Jackson and Chen, 2010). However, it is uncertain as to whether individual genes or WGD have contributed equally to the evolution and functional plasticity of plant genomes (see Dehal and Boore, 2005; De Storme and Mason, 2014). This ambiguity results in part because a direct causal link between an adaptive phenotype and a specific gene duplication event are difficult to ascertain because they usually occur at different times (Lawton-Rauth, 2003).

Studies during the past two decades have provided valuable information about intrinsically disordered/ductile proteins (IDPs) and disordered regions (Xie et al., 2007; Oldfield and Dunker, 2014; van der Lee et al., 2014; Wright and Dyson, 2015; Yan et al., 2016). IDPs do not fold into well-defined three-dimensional (3D) structures and can be either entirely disordered or partially disordered, with regions spanning just a few contiguous disordered residues (<10 aa) or containing long segments (≥ 30 aa) of contiguously disordered residues. Numerous researchers have developed algorithms that use amino acid sequences as inputs to predict the probability to be structured or disordered for each residue as outputs (He et al., 2009; Meng et al., 2017). By applying such disorder predictors to sequences of proteins with known functions, the biological activities of IDPs can be inferred from large collections of proteins (Ward et al., 2004; Xie et al., 2007). From these and other studies it has been concluded that 25–50% of all eukaryotic

proteins contain at least one long IDP region and that 33–50% of eukaryotic proteomes have IDPs regions.

At the molecular level, it is uncertain how disordered/ductile proteins evolve in the scenario of WGD and PPD events. Nevertheless, there is ample evidence that disordered residues (IDRs) confer flexibility to proteomes (Tomba, 2002; Schad et al., 2011; Yruela and Contreras-Moreira, 2012; Yruela et al., 2017). Moreover, IDPs are known (1) to have played a significant role in the evolution of multicellularity and/or cell type specification (Niklas et al., 2014, 2018; Dunker et al., 2015; Niklas and Dunker, 2016; Yruela et al., 2017), (2) to contribute to organismic plasticity by facilitating protein multifunctionalities and nucleic acid interactions through complex gene regulatory network dynamics (Dyson and Wright, 2002, 2005; Xie et al., 2007; Habchi et al., 2014; van der Lee et al., 2014; O'Shea et al., 2015; Wright and Dyson, 2015; Yruela, 2015; Covarrubias et al., 2017), and (3) to be associated with proteins involved in signaling, cellular regulation, nuclear localization, chaperone activity, and RNA, DNA, and protein binding among many other functions (Xie et al., 2007; Kovacs et al., 2008; Babu et al., 2011; Buljan et al., 2013; Pazos et al., 2013; Oldfield and Dunker, 2014; Skupien-Rabian et al., 2016; Olsen et al., 2017). Moreover, IDPs and IDRs collaborate with alternative splicing (AS) and post-transcriptional modifications (PTMs) to markedly enhance the complexity of signaling networks (Niklas et al., 2015). By means of these collaborations, the same gene products bring about alternative signaling outcomes that depend on the use by IDPs or IDRs of shape-shifts to bind to multiple, different partners and that depend on the further alteration of partner binding by AS and/or PTMs (Dunker et al., 2008, 2015; Rautureau et al., 2010; Hsu et al., 2013; Niklas et al., 2015; Zhou et al., 2018). This collaboration of IDPs or IDRs with AS and/or PTMs appears to have contributed significantly to the evolution of multicellularity in all major eukaryotic lineages (Niklas et al., 2018).

The goal of this paper is to evaluate the hypothesis that WGD (or WGT) events have disproportionately increased IDRs in plant proteomes thereby contributing to plant “evolvability.” To investigate this hypothesis, we determined the fraction of IDRs in co-linear paralogs of several model and economically important plant species including green algae, bryophytes, monocots, and eudicots.

MATERIALS AND METHODS

Protein Sequences

Protein sequences of co-linear paralogs of one chlorophyte (*Chlamydomonas reinhardtii*, $n = 32$), one bryophyte (*Physcomitrella patens*, $n = 3,716$), four monocots (*Zea mays*, $n = 14,062$; *Sorghum bicolor*, $n = 5,336$; *Oryza sativa*, $n = 6,503$; *Brachypodium distachyon*, $n = 4,670$), and thirteen eudicots (*Glycine max*, $n = 74,584$; *Populus trichocarpa*, $n = 27,976$; *Brassica oleracea*, $n = 41,318$; *Manihot esculenta*, $n = 18,954$; *Vitis vinifera*, $n = 8,836$; *Gossypium raimondi*, $n = 25,880$; *Capsicum annuum*, $n = 996$; *Solanum lycopersicum*, $n = 6,796$; *Arabidopsis thaliana*, $n = 7,894$; *Prunus persica*, $n = 4,784$; *Beta vulgaris*, $n = 774$; *Medicago truncatula*, $n = 6,664$; *Cucumis*

sativus, $n = 2,198$) were retrieved from PLAZA 4.0,¹ (van Bel et al., 2017). These species were selected because (1) some are model experimental systems (e.g., *C. reinhardtii*, *P. patens*, *A. thaliana*, *B. distachyon*), (2) others are economically extremely important (e.g., *Z. mays*, *G. max*, *V. vinifera*), and (3) all have full representative genome and chromosome assemblies. Co-linear regions within genomes are annotated in PLAZA 4.0 by application of the i-ADHoRe algorithm (Proost et al., 2012). Co-linear paralogs are encoded by genes from the same gene family and are located in genomic segments that share the same gene content in the same order.

WGD and WGT in Plants

Ancient large-scale duplication (WGD) and triplication (WGT) events, or more recent duplications have been reported in the literature for the following model plant systems: *P. patens* (1 recent duplication), *Z. mays* (6 duplications), *S. bicolor* (5 duplications), *O. sativa* (5 duplications), *B. distachyon* (5 duplications), *G. max* (4 duplications and 1 triplication), *P. trichocarpa* (3 duplications and 1 triplication), *V. vinifera* (2 duplications and 1 triplication), *B. oleracea* (2 duplications), *M. esculenta* (2 duplications and 1 triplication), *G. raimondi* (2 duplications and 1 undefined event), *S. lycopersicum* (2 duplications and 2 triplications), *A. thaliana* (5 duplications), *P. persica* (2 duplications and 1 triplication), *M. truncatula* (3 duplications and 1 triplication) (Panchy et al., 2016; for more details about WGD and WGT history see Gaut et al., 2000; Blanc and Wolfe, 2004; Yu et al., 2005; Schmutz et al., 2010; Song et al., 2012; Tiley et al., 2016).

Mean reported K_s values, which represent WGDs and the divergence of duplicate gene pairs in plant families, are as follows: Angiosperms ($K_s > 3$), *Solanaceae* ($K_s = 0.60$), *Fabaceae* ($K_s = 0.60$), *Poaceae* ($K_s = 0.90$), *Brassicaceae* ($K_s = 0.80$) (Schranz et al., 2012).

Prediction of Disordered Residues

More than 60 predictors of disorder have been developed (He et al., 2009). A comparison of predictors and their variants across 1,765 proteomes reveals considerable variation in their ability to identify IDRs (Oates et al., 2013), indicating that the reliability of the predictor used in this study had to be evaluated critically before detailed analyses were undertaken. In a detailed comparison of 16 commonly used predictors, PONDR VSL2b (Peng et al., 2006) had the best overall accuracy for long disordered regions (Peng and Kurgan, 2012). Nevertheless, we also explored DISOPRED v3.1 (Jones and Cozzetto, 2015) using a selected group of model monocot and eudicot species. Our analyses showed that overall DISOPRED v3.1 provided consistent results with the predictions of PONDR VSL2b ($r^2 = 0.60$ and $r^2 = 0.96$ for predicted IDRs and disordered regions with $L > 30$ aa, respectively) (Supplementary Figures S1–S3).

Data Analysis

Bar-plots and statistical analysis were performed with Origin Pro8.6. The coefficient of determination, r^2 , of standard linear

regression protocols were calculated as:

$$r^2 = 1 - \left(\frac{\text{RSS}/\text{df}_{\{\text{Error}\}}}{\text{TSS}/\text{df}_{\{\text{Error}\}}} \right)$$

where RSS is the residual sum of square and TSS is the total sum of square (Anderson-Sprecher, 1994).

Sequence Alignment

Pairwise alignments of co-linear paralogous sequences were determined using Clustal Omega (Sievers et al., 2011). For each aligned pair, the aligned disorder predictions were compared in order to calculate three types of IDRs: (1) identical disordered residues, where both the amino acid sequence and disorder predictions were identical (denoted hereafter as “identical IDRs”), (2) similar disordered residues, where the disorder predictions matched but the amino acid sequence varied (“similar IDRs”), and (3) variable residues where disorder predictions were not conserved (“variable IDRs”) (Yruea et al., 2017). The same three IDRs types were also computed for the subset of residues that were predicted to be disordered within long segments of at least 30 contiguous disordered residues ($L > 30$ aa). In all cases the fraction of IDRs were computed by dividing the number of aligned IDRs with the total aligned residues.

The IDR categories described previously (Yruea et al., 2017) and used here were inspired by work of Bellay et al. (2011) but differ in some details reported previously by others. Bellay et al. (2011) focused on orthologs and did not consider insertions and deletions, only sequences that could be aligned. In contrast, here we are studying paralogs. The three categories of IDRs described above for our work provided useful categories for the IDRs found in these proteins. For such proteins, we were particularly interested in examples in which disordered/ductile regions were present and absent in a given paralogous pair of proteins, and as noted above, Bellay et al. (2011), did not consider insertions or deletions at all.

Gene Ontology (GO) Enrichment Analysis

Gene Ontology annotations for the complete proteomes analyzed were retrieved from PLAZA 4.0 and genome GO term (expected) frequencies computed for each species. In addition, the subset of GO annotations corresponding to pairs of paralogs harboring long disordered segments ($L > 30$ aa) were used to compute (1) observed GO term frequencies for co-linear genes and (2) observed GO term frequencies for co-linear genes harboring at least two thirds of identical IDRs. Enrichment was computed by applying Fisher's exact test with Bonferroni correction² to compare the observed and expected GO term frequencies. When possible, plant-specific GO-slim terms were assigned to enriched terms by parsing file³.

RESULTS AND DISCUSSION

We focused on the paralogs of 19 important plant species across a broad spectrum of the Chlorobionta (i.e., green algae to eudicots)

²<http://stat.ethz.ch/R-manual/R-devel/library/stats/html/fisher.test.html>

³http://www.geneontology.org/ontology/subsets/goslim_plant.obo

whose proteome size, basic haploid chromosome number and number of co-linear paralogous pairs differ significantly (Table 1). From 37 to 52% total aligned paralogous sequences were identified by PONDR VSL2b as having an IDR signature. These results are consistent with previous whole proteome analyses (Yruela and Contreras-Moreira, 2012). The highest percentages of aligned IDRs were found in monocot species (50–52%); the lowest percentage was found in the green alga *C. reinhardtii* (37%). The range of total aligned IDRs observed for the 19 species examined in this study accords reasonably well with the evolutionary origins of these taxa (i.e., total aligned IDRs tend to increase with more recent descent). It is worth noting that these values were on average much lower than those predicted using DisoPred v3.1 (Supplementary Table S1) and also much lower than those previously reported (Yruela and Contreras-Moreira, 2012) using DisoPred v2.42 (Ward et al., 2004). The differences observed using both versions of DisoPred are attributed to different sensitivities to IDRs longer than 20 amino acids (Jones and Cozzetto, 2015).

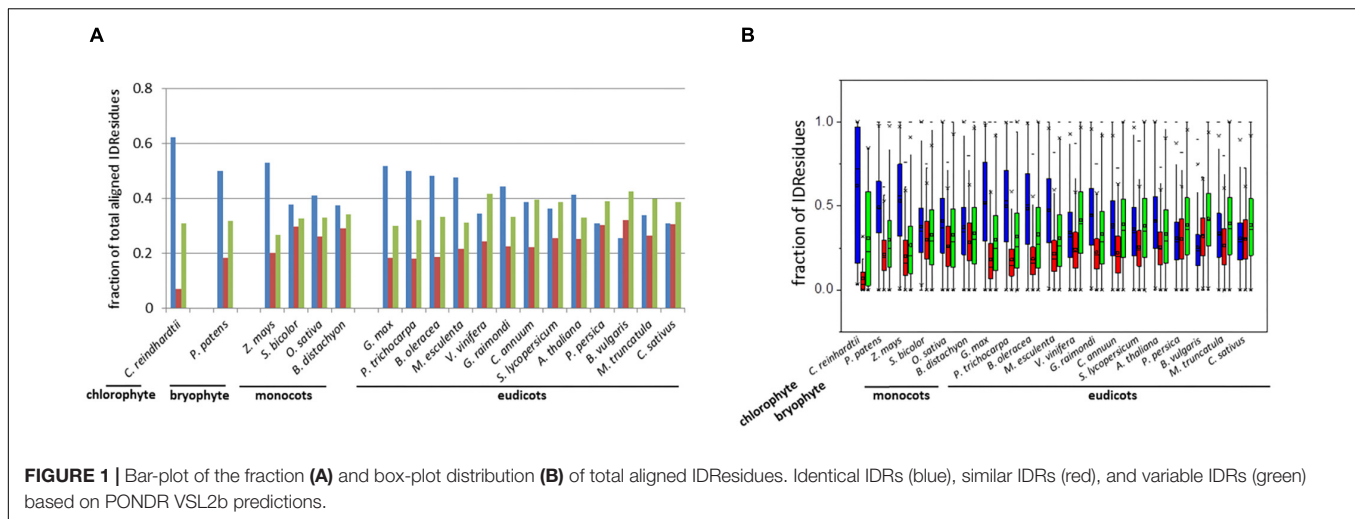
The fraction of aligned residues in long disordered segments ($L > 30$ aa) was also calculated using PONDR VSL2b, which revealed high proportions of IDRs located in such segments $L > 30$ aa (ca. 60%) (Table 1). Similar results were obtained using DisoPred v3.1 (Supplementary Table S1). Thus, the results consistently indicated that a high proportion of IDRs reside in long ductile segments ($L > 30$ aa) in paralogous pairs.

The distribution of the three types of IDRs (i.e., identical, similar, and variable IDRs) was analyzed for the paralogous pairs across all 19 species (Figure 1). Here we use the terms “identical IDRs” for those that are conserved with respect to sequence, length and location from one paralog to the next, “similar IDRs” for those that show substantial sequence variations but are conserved with respect to length and location from one paralog to the next and “variable IDRs” for those that are observed in some paralogs but absent in others.

Analyses indicated that the percentage of aligned identical IDRs in paralogous sequences predicted by PONDR VSL2b ranged between 30 and 60%. It was highest in the green algae *C. reinhardtii* and lowest in the eudicot *B. vulgaris* (Figure 1). These data are in agreement with previous results (Yruela and Contreras-Moreira, 2013). The predicted fractions of similar IDRs and variable IDRs were highest in *B. vulgaris*, *M. truncatula*, *P. persica*, and *C. sativus* (Figures 1, 2). We speculate that the differences observed among the three different kinds of predicted IDRs reflect the history of genome duplication/polyploidy events (i.e., both chromosome number and the number of paralogs) in the species investigated in this study (Table 1). It is worth noting that the basic haploid chromosome number of *B. vulgaris*, *M. truncatula*, *P. persica*, and *C. sativus* are much reduced ($n = 7 - 9$) compared with those of the green alga *C. reinhardtii* ($n = 17$), other monocots such as *O. sativa* ($n = 12$) and *Z. mays* ($n = 10$), and eudicots such as *G. max* ($n = 20$) and *P. trichocarpa* ($n = 19$).

TABLE 1 | Characteristics of the nineteen plant species examined in this study.

Plant species	Proteome size	Haploid chromosome number (n)	Pair of co-linear paralogs	Fraction of total aligned IDRs PONDR VSL2	Fraction of aligned IDRs in segments $L > 30$ aa PONDR VSL2b
Chlorophyte (n)					
<i>Chlamydomonas reinhardtii</i>	14,488	17	16	0.37	0.55
Bryophyte					
<i>Physcomitrella patens</i>	48,022	27	1,858	0.50	0.62
Monocots ($2n = 2 \times$)					
<i>Zea mays</i>	58,615	10	7,031	0.52	0.66
<i>Sorghum bicolor</i>	39,248	10	2,668	0.52	0.65
<i>Oryza sativa</i>	40,881	12	3,251	0.50	0.62
<i>Brachypodium distachyon</i>	33,844	5	2,335	0.50	0.63
Eudicots ($2n = 2 \times$)					
<i>Glycine max</i>	71,525	20	37,292	0.45	0.56
<i>Populus trichocarpa</i>	73,110	19	13,988	0.46	0.55
<i>Vitis vinifera</i>	41,208	19	1,916	0.44	0.53
<i>Brassica oleracea</i>	56,687	9	20,660	0.48	0.57
<i>Manihot esculenta</i>	43,286	18	9,477	0.47	0.58
<i>Gossypium raimondi</i>	59,057	13	12,941	0.47	0.58
<i>Capsicum annuum</i>	40,627	12	498	0.42	0.48
<i>Solanum lycopersicum</i>	36,010	12	3,399	0.46	0.55
<i>Arabidopsis thaliana</i>	48,148	5	3,947	0.47	0.59
<i>Prunus persica</i>	32,595	8	2,392	0.47	0.58
<i>Beta vulgaris</i>	32,874	9	387	0.47	0.57
<i>Medicago truncatula</i>	57,661	8	3,322	0.47	0.57
<i>Cucumis sativus</i>	25,668	7	1,099	0.49	0.59



(Table 1). Furthermore, the combination of multiple ancestral WGD and more recent polyploidy events promoting high rates of duplicated gene retentions (e.g., *P. trichocarpa*, *G. max*, *B. oleracea*) (Parkin et al., 2014; Panchy et al., 2016) likely also favored the increase of identical IDRs.

With the exception of the green alga *C. reinhardtii* and the bryophyte *P. patens*, a statistically significant and positive correlation ($r^2 = 0.45$, $P = 8 \times 10^{-4}$) was observed between the number of co-linear paralogous pairs and the haploid number of chromosomes across the 17 vascular plant species (Supplementary Figure S2A). The proteome size and the number of paralogs were also significantly correlated with one another ($r^2 = 0.53$, $P = 5 \times 10^{-3}$) (Supplementary Figure S2B).

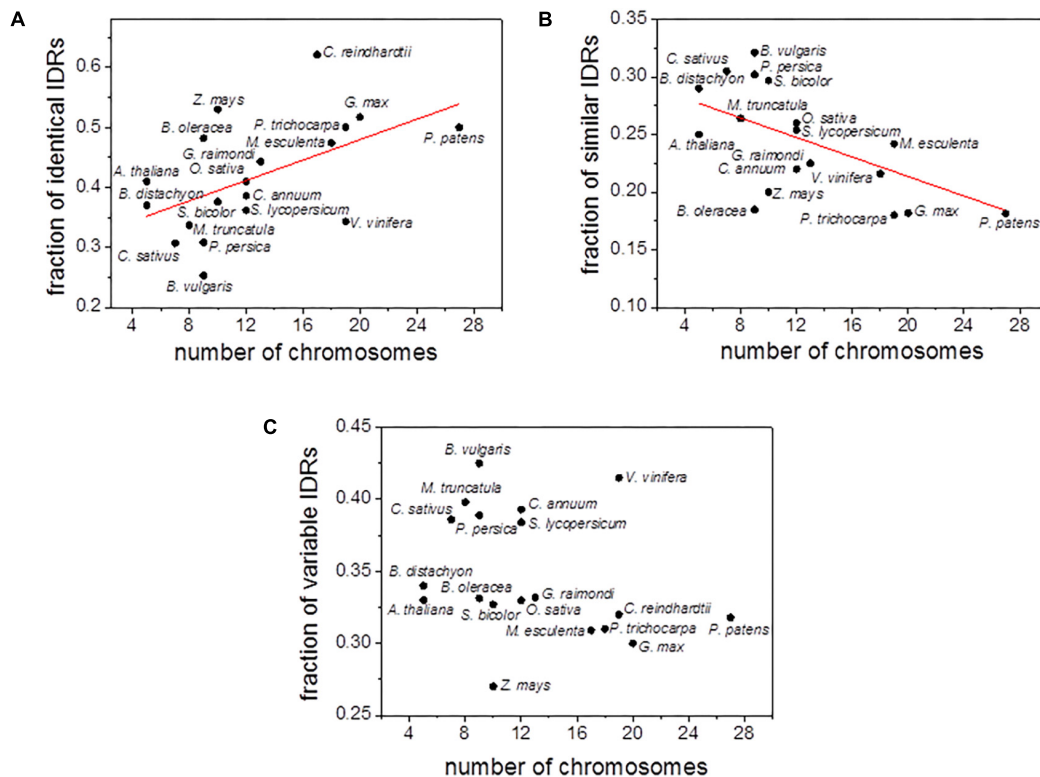
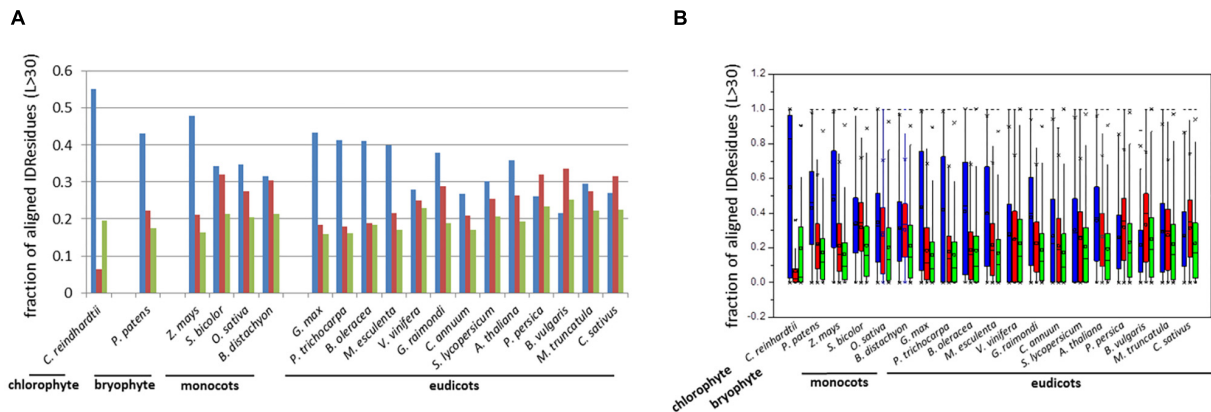
In order to further explore the relationship between polyploidy and IDRs content, we analyzed the correlation between the number of chromosomes and the fraction of the three types of IDRs (i.e., identical, similar and variable IDRs). A statistically positive and significant correlation ($r^2 = 0.42$, $P = 5 \times 10^{-3}$) was observed between the number of chromosomes and the fraction of identical IDRs (Figure 3A and Supplementary Figure S3A). In contrast, a statistically significant negative correlation ($r^2 = 0.42$, $P = 5 \times 10^{-3}$) was observed for the fraction of similar IDRs (Figure 3B and Supplementary Figure S3B). Little or no correlation was observed between the number of chromosomes and the fraction of variable IDRs (Figure 3C and Supplementary Figure S3C).

It has been reported that most of the retained duplicated genes in angiosperms are enriched in Gene Ontology (GO) categories associated with protein targeting, synthesis, and post-translational modification (Ren et al., 2018). In order to put in perspective our results and get additional insights, we investigated the GO annotations function of (1) co-linear paralogous proteins in all 19 plant species studied, and (2) the group of co-linear paralogous harboring a majority of identical IDRs. The analysis revealed that on average paralogs are enriched in biological processes (P) (50–60%), molecular functions (F) (20–30%) and cellular components (C) (15–30%) GO categories with corrected p -values $< 10E-6$. Similar trends

were found in the group of paralogs enriched in identical IDRs (p -values $< 10E-5$). Regarding biological processes, we found that paralogs with identical IDRs are mainly associated with terms such as “catalytic activity,” “metabolic process,” “biosynthetic process,” “development,” “cell differentiation” and “cell proliferation” (p -values $< 10E-6$). The most significant association among specific molecular functions was with “molecular binding” and “transport” terms (p -values $< 10E-6$). Regarding cellular components we found that paralogs with identical IDRs are associated with “plasma membrane” and “thylakoid” terms (p -values $< 10E-6$).

Differences in the distribution of the fraction of IDRs across the co-linear paralogous sequences could be the result of differences in the locations of paralogous genes along chromosomes. This attribution is based on a positive correlation between genetic recombination rates and protein disorder frequency observation, and on the fact that ductile segments are more conserved between paralogs located in regions close to (as opposed to distant from) centromeres (Yruea and Contreras-Moreira, 2013). It is clear from previous analyses and the results presented here that significant evolutionary differences exist in proteomes and in the “dynamics” of IDRs protein sequence sorting during polyploidy events, i.e., our data indicate that polyploidy incurs a disproportionate increase in highly conserved flexibility/ductility compared with less conserved and random disordered/ductile protein regions.

Differences in the distribution of the three types of IDRs have been also observed in a set of transcription factor orthologs involved in key developmental processes such as cellular differentiation, cell division, cell cycle, and cell proliferation (Yruea et al., 2017). Analyses indicated that the fraction of predicted aligned identical IDRs is higher in the green algae (chlorophytes) and non-vascular land plants (bryophytes) compared to vascular plants and animals, whereas the fraction of less conserved IDRs (similar and variable IDRs) is lower in the green algae (chlorophytes) and the non-vascular plants in comparison to vascular plants and animals (Yruea et al., 2017).



Inspection of **Figure 4A** shows that co-linear paralogs of GATA10 (AT1G08000), located on chromosome 2 (AT2G28340) and chromosome 3 (AT3G54810), have important differences in the distribution of IDRs, as indicated in the marked zinc-finger GATA-type binding domain. Although the three transcription factors have a high proportion of IDRs (ca. 90%), analysis

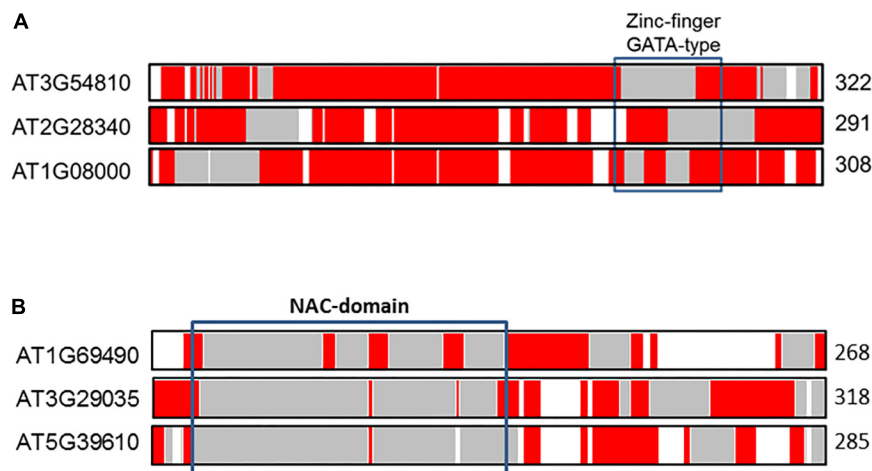


FIGURE 4 | Alignments of the co-linear GATA10 **(A)** and NAC92 **(B)** paralogs in *A. thaliana*. The protein length is given to the right of each sequence. The color coded bars in each sequence depict predicted disorder. Disordered residues are shown in red, ordered residues are shown in gray, and alignment gaps are shown in white. Typical DNA-binding domains are shown as black boxes.

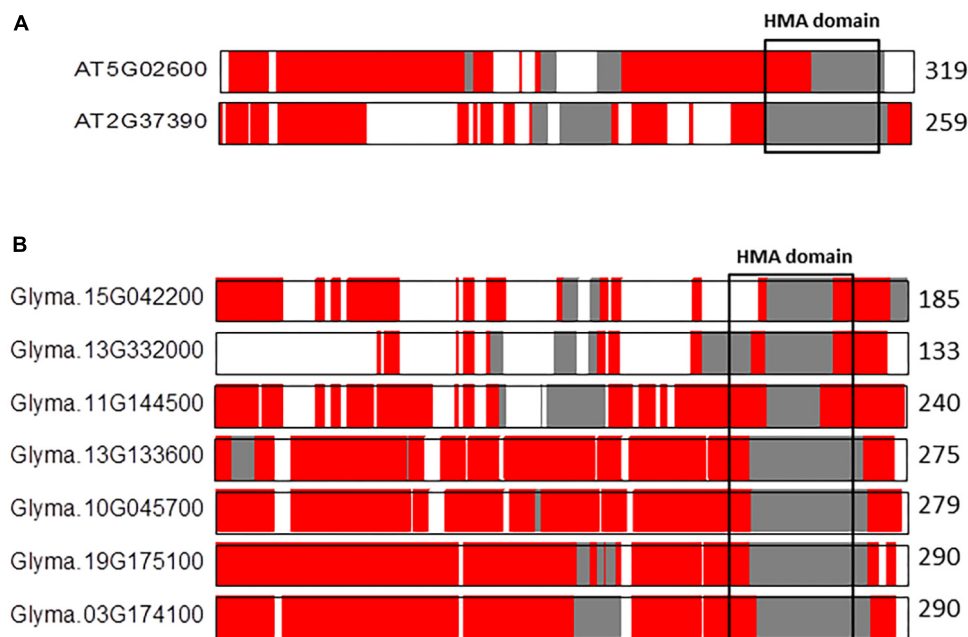


FIGURE 5 | Alignments of the co-linear NAKR1 paralogs in *A. thaliana* **(A)** and *G. max* **(B)**. The protein length is given to the right of each sequence. Sequences are represented by color coded bars representing predicted disorder. Disordered residues are shown in red, ordered residues are shown in gray, and alignment gaps are shown in white. Typical DNA-binding domains are shown as black boxes.

indicates that between 13 and 30% of the aligned residues correspond to identical IDRs. The percentage of similar and variable IDRs ranges from between 30 and 46%. It has been speculated that the three paralogs are involved in cell differentiation, and that they might be involved in the regulation of some light-responsive genes. We speculate further that variations observed in the distribution of IDRs around the DNA-binding motif might result in different paralog functionalities. Such differences contrast with those observed in GATA orthologs

(Yruela et al., 2017). In particular the distribution of IDRs in the zinc-finger GATA-type binding domain is more conserved and manifests a progressive gain of IDRs from green algae to vascular plants, which increases flexibility/ductility in the functional domain.

The alignment of NAC92 (AT5G39610) paralogs on chromosome 3 (AT3G29035 or NAC59) and chromosome 1 (AT1G69490 or NAC29) also reveals notable differences in IDRs distributions. The percentage of total aligned IDRs is ca.

40%, and that of identical IDRs is ca. 10–20%. Such differences likely contributed to functional divergences. NAC92 and NAC59 are involved in senescence, salt stress responses, and lateral root development (Balazadeh et al., 2010), whereas NAC29 is involved in heat stress responses (Zhao et al., 2018).

An additional interesting example is the comparison between the co-linear NAKR paralogs in *A. thaliana* ($n = 5$) and *G. max* ($n = 20$) (Figure 5). The pair of co-linear NAKR1 (AT5G02600) and NAKR2 (AT2G37390) paralogs located on *A. thaliana* chromosomes 5 and 3, respectively, show notable differences in the IDRs distribution along sequences (Figure 5A). The percentage of total aligned IDRs is ca. 82% in contrast to the percentage of identical IDRs, which is only ca. 24%. The percentage of similar and variable IDRs is 52 and 22%, respectively. The alignment reveals once again differences in the distribution of IDRs, particularly in the functional HMA domain. NAKR1 (Sodium Potassium Root Defective1) is a heavy metal-binding protein expressed in phloem. It interacts with the FLOWERING LOCUS T (FT) transcription factor and regulates flowering through both the transcriptional regulation and transport of FT, especially in response to potassium availability (Negishi et al., 2018). The precise function of NAKR2 is still unclear. In contrast, in *G. max* the differences in the IDRs composition of the HMA domain across the seven co-linear NAKR1 paralogs are smaller, in particular among four of them, Glyma13G133600, Glyma10G045700, Glyma19G175100, and Glyma03G174100 located on *G. max* chromosomes 13, 10, 19, and 3, respectively (Figure 5B). The proportion of aligned IDRs on average ranges from 53 to 81%. The fraction of identical IDRs is ca. 87%, and those of similar IDRs and variable IDRs are ca. 1 and 0.6%, respectively. These observations once again support our hypothesis that polyploidy likely favors increases in highly conserved flexibility/ductility. This fact might have preserved essential functionalities during the course of angiosperm evolution.

CONCLUSION

In summary, the results reported here indicate: (1) a positive correlation between chromosome number and the fraction of paralogous sequence that are identified as identical IDRs, and (2) a negative correlation between chromosome number and the fraction of paralogous sequences that are identified as similar IDRs. We interpret these findings to indicate (1) retention of paralogs with identical IDRs after WGD (or WTG) could be favored by selection because identical IDRs (as opposed to similar/variable IDRs) facilitated essential functions involved in development, and (2) the retention of genes with high proportions of similar/variable IDRs after WGD (or WTG) could be less likely and therefore tended to be lost in one of paralogs. We argue that the patterns observed

for similar/variable IDRs pattern are simply a byproduct of recent WGD (or WTG) events. Thus, ancient WGD (or WGT) events in species such as *Z. mays*, *G. max*, and *P. trichocarpa* have disproportionately favored an increase in aligned identical IDRs across paralogs, thereby contributing to the stability of functions such as the catalytic activity of proteins, metabolic and transport processes, and molecular binding. Based on these characteristics, it is not unreasonable to speculate that, over evolutionary time, duplication events have stabilized proteome adaptive functionalities.

AUTHOR CONTRIBUTIONS

IY and AKD conceived the study. IY analyzed the data, wrote the original draft of the manuscript, and reviewed and edited the manuscript. BC-M did data analyses, and reviewed and edited the manuscript. AKD and KN contributed to discussion, and wrote, reviewed, and edited the manuscript.

FUNDING

This work was supported by Gobierno de Aragón (DGA-GC E35_17R and A08_17R). These grants were partially financed by the EU FEDER Program. We acknowledge support of the publication fee by the CSIC Open Access Publication Support Initiative through its Unit of Information Resources for Research (URICI).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.01216/full#supplementary-material>

FIGURE S1 | Average fraction (A,B) and box-plot distribution (C,D) of total aligned IDRs (A,B) and aligned residues in ductile regions ($L > 30$ aa). Identical IDRs (blue), similar IDRs (red), and variable IDRs (green). The data represent the average of paralogs in the proteomes of three monocots (*Z. mays*, *O. sativa*, *B. distachyon*) and two eudicots (*P. trichocarpa*, *A. thaliana*). Disordered predictions are based on DisoPred v3.1.

FIGURE S2 | Scatter plots of (A) number of chromosomes versus number of paralogous pairs and (B) proteome size versus number of paralogous pairs in monocots (*Z. mays*, *O. sativa*, *B. distachyon*) and eudicots (*P. trichocarpa*, *A. thaliana*). Disordered predictions are based on DisoPred v3.1.

FIGURE S3 | Scatter plots of number of chromosomes versus the fraction of aligned identical IDRs (A), similar IDRs (B), and variable IDRs (C) in monocots *Z. mays* ($n = 10$), *O. sativa* ($n = 12$), *B. distachyon* ($n = 5$), and eudicots *P. trichocarpa* ($n = 19$) and *A. thaliana* ($n = 5$). Disordered predictions are based on DisoPred v3.1.

TABLE S1 | Characteristics of the plant species examined with DisoPred v3.1.

REFERENCES

Anderson-Sprecher, R. (1994). Model comparisons and R2. *Am. Stat.* 48, 113–117.

Babu, M. M., van der Lee, R., de Groot, N. S., and Gsponer, J. (2011). Intrinsically disordered proteins: regulation and disease. *Curr. Opin. Struct. Biol.* 21, 432–440. doi: 10.1016/j.sbi.2011.03.011

- Balazadeh, S., Siddiqui, H., Allu, A. D., Matallana-Ramirez, L. P., Caldana, C., Mehrnia, M., et al. (2010). A gene regulatory network controlled by the NAC transcription factor ANAC092/AtNAC2/ORE1 during salt-promoted senescence. *Plant J.* 62, 250–264. doi: 10.1111/j.1365-313X.2010.04151.x
- Bellay, J., Han, S., Michaut, M., Kim, T., Costanzo, M., Andrews, B. J., et al. (2011). Bringing order to protein disorder through comparative genomics and genetic interactions. *Genome Biol.* 12:R14. doi: 10.1186/gb-2011-12-2-r14
- Blanc, G., and Wolfe, K. H. (2004). Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* 16, 1667–1678. doi: 10.1105/tpc.021345
- Buljan, M., Chalancon, G., Dunker, A. K., Bateman, A., Balaji, S., Fuxreiter, M., et al. (2013). Alternative splicing of intrinsically disordered regions and rewiring of protein interactions. *Curr. Opin. Struct. Biol.* 23, 443–450. doi: 10.1016/j.sbi.2013.03.006
- Covarrubias, A. A., Cuevas-Velazquez, C. L., Romero-Pérez, P. S., Rendón-Luna, D. F., and Chater, C. C. C. (2017). Structural disorder in plant proteins: where plasticity meets sessility. *Cell Mol. Life Sci.* 74, 3119–3147. doi: 10.1007/s00018-017-2557-2
- De Storme, N., and Mason, A. (2014). Plant speciation through chromosome instability and ploidy change: cellular mechanisms, molecular factors and evolutionary relevance. *Curr. Plant Biol.* 1, 10–33. doi: 10.1016/j.cpb.2014.09.002
- Dehal, P., and Boore, J. L. (2005). Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol.* 3:e314. doi: 10.1371/journal.pbio.0030314
- del Pozo, J. C., and Ramírez-Parra, E. (2015). Whole genome duplications in plants: an overview from Arabidopsis. *J. Exp. Bot.* 66, 6991–7003. doi: 10.1093/jxb/erv432
- Dubcovsky, J., and Dvorak, J. (2007). Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science* 316, 1862–1866. doi: 10.1126/science.1143986
- Dunker, A. K., Bondos, S. E., Huang, F., and Oldfield, C. J. (2015). Intrinsically disordered proteins and multicellular organisms. *Semin. Cell Dev. Biol.* 37, 44–55. doi: 10.1016/j.semcdb.2014.09.025
- Dunker, A. K., Silman, I., Uversky, V. N., and Sussman, J. L. (2008). Function and structure of inherently disordered proteins. *Curr. Opin. Struct. Biol.* 18, 756–764. doi: 10.1016/j.sbi.2008.10.002
- Dyson, H. J., and Wright, P. E. (2002). Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* 12, 54–60. doi: 10.1016/S0959-440X(02)00289-0
- Dyson, H. J., and Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* 6, 197–208. doi: 10.1038/nrm1589
- Gaut, B. S., Le Thierry d'Ennequin, M., Peek, A. S., and Sawkins, M. C. (2000). Maize as a model for the evolution of plant nuclear genomes. *Proc. Natl. Acad. Sci. U.S.A.* 97, 7008–7015. doi: 10.1073/pnas.97.13.7008
- Habchi, J., Tompa, P., Longhi, S., and Uversky, V. N. (2014). Introducing protein intrinsic disorder. *Chem. Rev.* 114, 6561–6588. doi: 10.1021/cr400514h
- He, B., Wang, K., Liu, Y., Xue, B., Uversky, V. N., and Dunker, A. K. (2009). Predicting intrinsic disorder in proteins: an overview. *Cell Res.* 19, 929–949. doi: 10.1038/cr.2009.87
- Hsu, W. L., Oldfield, C. J., Xue, B., Meng, J., Huang, F., Romero, P., et al. (2013). Exploring the binding diversity of intrinsically disordered proteins involved in one-to-many binding. *Protein Sci.* 22, 258–273. doi: 10.1002/pro.2207
- Jackson, S., and Chen, Z. J. (2010). Genomic and expression plasticity of polyploidy. *Curr. Opin. Plant Biol.* 13, 153–159. doi: 10.1016/j.pbi.2009.11.004
- Jones, D. T., and Cozzetto, D. (2015). DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics* 31, 857–863. doi: 10.1093/bioinformatics/btu744
- Kagale, S., Robinson, S. J., Nixon, J., Xiao, R., Huebert, T., Condie, J., et al. (2014). Polyploid evolution of the Brassicaceae during the Cenozoic era. *Plant Cell* 26, 2777–2791. doi: 10.1105/tpc.114.126391
- Kellogg, E. A., and Bennetzen, J. L. (2004). The evolution of nuclear genome structure in seed plants I. *Am. J. Bot.* 91, 1709–1725. doi: 10.3732/ajb.91.10.1709
- Kovacs, D., Kalmar, E., Torok, Z., and Tompa, P. (2008). Chaperone activity of ERD10 and ERD14, two disordered stress-related plant proteins. *Plant Physiol.* 147, 381–390. doi: 10.1104/pp.108.118208
- Landis, J. B., Soltis, D. E., Li, Z., Marx, H. E., Barker, M. S., Tank, D. C., et al. (2018). Impact of whole-genome duplication events on diversification rates in angiosperms. *Am. J. Bot.* 2018:2. doi: 10.1002/ajb2.1060
- Lawton-Rauth, A. (2003). Evolutionary dynamics of duplicate genes in plants. *Mol. Phylogenet. Evol.* 29, 396–409. doi: 10.1016/j.ympev.2003.07.004
- Leitch, A. R., and Leitch, I. J. (2008). Genomic plasticity and the diversity of polyploid plants. *Science* 320, 481–483. doi: 10.1126/science.1153585
- Li, Z., Tiley, G. P., Galuska, S. R., Reardon, C. R., Kidder, T. I., Rundell, R. J., et al. (2018). Multiple large-scale gene and genome duplications during the evolution of hexapods. *Proc. Natl. Acad. Sci. U.S.A.* 115, 4713–4718. doi: 10.1073/pnas.1710791115
- Lyons, E., Pedersen, B., Kane, J., Alam, M., Ming, R., Tang, H., et al. (2008). Finding and comparing syntenic regions among *Arabidopsis* and the outgroups papaya, poplar and grape: CoGe with rosids. *Plant Physiol.* 148, 1772–1781. doi: 10.1104/pp.108.124867
- Mandáková, T., and Lysak, M. A. (2018). Post-polyploid diploidization and diversification through dysploid changes. *Curr. Opin. Plant Biol.* 42, 55–65. doi: 10.1016/j.pbi.2018.03.001
- Meng, F., Uversky, V., and Kurgan, L. (2017). Computational prediction of intrinsic disorder in proteins. *Curr. Protoc. Protein Sci.* 88, 2.16.1–2.16.14. doi: 10.1002/cpps.28
- Negishi, K., Endo, M., Abe, M., and Araki, T. (2018). SODIUM POTASSIUM ROOT DEFECTIVE1 regulates FLOWERING LOCUS T expression via the microRNA156-SQUAMOSA PROMOTER BINDING PROTEIN-LIKE3 module in response to potassium conditions. *Plant Cell Physiol.* 59, 404–413. doi: 10.1093/pcp/pcx199
- Niklas, K. J., Bondos, S. E., Dunker, A. K., and Newman, S. A. (2015). Rethinking gene regulatory networks in light of alternative splicing, intrinsically disordered protein domains, and post-translational modifications. *Front. Cell Dev. Biol.* 3:8. doi: 10.3389/fcell.2015.00008
- Niklas, K. J., Cobb, E. D., and Dunker, A. K. (2014). The number of cell types, information content, and the evolution of complex multicellularity. *Acta Soc. Bot. Pol.* 83, 337–347. doi: 10.5586/asbp.2014.034
- Niklas, K. J., and Dunker, A. K. (2016). “Alternative splicing, intrinsically disordered proteins, calmodulin, and the evolution of multicellularity,” in *Multicellularity: Origins and Evolution*, eds K. J. Niklas and S. A. Newman (Cambridge, MA: MIT Press), 17–40.
- Niklas, K. J., Dunker, A. K., and Yruela, I. (2018). The evolutionary origins of cell type diversification and the role of intrinsically disordered proteins. *J. Exp. Bot.* 69, 1437–1446. doi: 10.1093/jxb/erx493
- Oates, M. E., Romero, P., Ishida, T., Ghalwash, M., Mizianty, M. J., Xue, B., et al. (2013). D²P²: database of disordered protein predictions. *Nucleic Acids Res.* 41, D508–D516. doi: 10.1093/nar/gks1226
- Oldfield, C. J., and Dunker, A. K. (2014). Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu. Rev. Biochem.* 83, 553–584. doi: 10.1146/annurev-biochem-072711-164947
- Olsen, J. G., Teilum, K., and Kragelund, B. B. (2017). Behaviour of intrinsically disordered proteins in protein-protein complexes with an emphasis on fuzziness. *Cell Mol. Life Sci.* 74, 3175–3183. doi: 10.1007/s00018-017-2560-7
- O'Shea, C., Jensen, M. K., Stender, E. G. P., Kragelund, B. B., Willemoes, A., and Skriver, K. (2015). Protein intrinsic disorder in *Arabidopsis* NAC transcription factors: transcriptional activation by ANAC013 and ANAC046 and their interactions with RCD1. *Biochem. J.* 465, 281–294. doi: 10.1042/BJ20141045
- Panchy, N., Lehti-Shiu, M., and Shiu, S.-H. (2016). Evolution of gene duplication in plants. *Plant Physiol.* 171, 2294–2316. doi: 10.1104/pp.16.00523
- Panopoulou, G., Hennig, S., Groth, D., Krause, A., Poustka, A. J., Herwig, R., et al. (2003). New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal genomes. *Genome Res.* 13, 1056–1066. doi: 10.1101/gr.874803
- Parkin, I. A., Koh, C., Tang, H., Robinson, S. J., Kagale, S., Clarke, W. E., et al. (2014). Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid *Brassica oleracea*. *Genome Biol.* 15:R77. doi: 10.1186/gb-2014-15-6-r77
- Pazos, F., Pietrosemoli, N., García-Martin, J. A., and Solano, R. (2013). Protein intrinsic disorder in plants. *Front. Plant Sci.* 4:363. doi: 10.3389/fpls.2013.00363
- Peng, K., Radivojac, P., Vucetic, S., Dunker, A. K., and Obradovic, Z. (2006). Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics* 7:208. doi: 10.1186/1471-2105-7-208
- Peng, Z.-L., and Kurgan, L. (2012). Comprehensive comparative assessment of in-silico predictors of disordered regions. *Curr. Protein Pept. Sci.* 13, 6–18. doi: 10.2174/138920312799277938

- Proost, S., Fostier, J., Witte, D., Dhoedt, B., Demeester, P., Van de Peer, Y., et al. (2012). i-ADHoRe 3.0—fast and sensitive detection of genomic homology in extremely large data sets. *Nucleic Acids Res.* 40:e11. doi: 10.1093/nar/gkr955
- Ramsey, J., and Schemske, D. W. (2002). Neopolyploidy in flowering plants. *Annu. Rev. Ecol. Syst.* 33, 589–639. doi: 10.1146/annurev.ecolsys.33.010802.150437
- Rautureau, G. J., Day, C. L., and Hinds, M. G. (2010). Intrinsically disordered proteins in bcl-2 regulated apoptosis. *Int. J. Mol. Sci.* 11, 1808–1824. doi: 10.3390/ijms11041808
- Ren, R., Wang, H., Guo, C., Zhang, N., Zeng, L., Chen, Y., et al. (2018). Widespread whole genome duplications contribute to genome complexity and species diversity in angiosperms. *Mol. Plant.* 11, 414–428. doi: 10.1016/j.molp.2018.01.002
- Renny-Byfield, S., and Wendel, J. F. (2014). Doubling down on genomes: polyploidy and crop plants. *Am. J. Bot.* 101, 1711–1725. doi: 10.3732/ajb.1400119
- Schad, E., Tompa, P., and Hegyi, H. (2011). The relationship between proteome size, structural disorder and organism complexity. *Genome Biol.* 12:R120. doi: 10.1186/gb-2011-12-12-r120
- Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., et al. (2010). Genome sequence of the palaeopolyploid soybean. *Nature* 463, 178–183. doi: 10.1038/nature08670
- Schranz, M. E., Mohammad, S., and Edger, P. P. (2012). Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model. *Curr. Opin. Plant Biol.* 15, 147–153. doi: 10.1016/j.pbi.2012.03.011
- Sievers, F., Wilm, A., Dineen, D. G., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7:539. doi: 10.1038/msb.2011.75
- Skupien-Rabian, B., Jankowska, U., Swiderska, B., Lukasiewicz, S., Ryszawy, D., Dziedzicka-Wasylewska, M., et al. (2016). Proteomic and bioinformatics analysis of a nuclear intrinsically disordered proteome. *J. Proteomics* 130, 76–84. doi: 10.1016/j.jprot.2015.09.004
- Soltis, P. S., and Soltis, D. E. (2009). The role of hybridization in plant speciation. *Annu. Rev. Plant Biol.* 60, 561–588. doi: 10.1146/annurev.arplant.043008.092039
- Song, C., Guo, J., Sun, W., and Wang, Y. (2012). Whole genome duplication of intra- and inter-chromosomes in the tomato genome. *J. Genet. Genomics* 39, 361–368. doi: 10.1016/j.jgg.2012.06.002
- Song, Q., and Chen, Z. J. (2015). Epigenetic and developmental regulation in plant polyploids. *Curr. Opin. Plant Biol.* 24, 101–109. doi: 10.1016/j.pbi.2015.02.007
- Tiley, G. P., Ané, C., and Burleigh, J. G. (2016). Evaluating and characterizing ancient whole-genome duplications in plants with gene count data. *Genome Biol. Evol.* 8, 1023–1037. doi: 10.1093/gbe/evw058
- Tompa, P. (2002). Intrinsically unstructured proteins. *Trends Biochem. Sci.* 27, 527–533. doi: 10.1016/S0968-0004(02)02169-2
- van Bel, M., Diels, T., Vancaester, E., Kreft, L., Botzki, A., van de Peer, Y., et al. (2017). PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Res.* 46, D1190–D1196. doi: 10.1093/nar/gkx1002
- van der Lee, R., Buljan, M., Lang, B., Weatheritt, R. J., Daughdrill, G. W., Dunker, A. K., et al. (2014). Classification of intrinsically disordered regions and proteins. *Chem. Rev.* 114, 6589–6631. doi: 10.1021/cr400525m
- Ward, J. J., McGuffin, L. J., Bryson, K., Buxton, B. F., and Jones, D. T. (2004). The DISOPRED server for the prediction of protein disorder. *Bioinformatics* 20, 2138–2139. doi: 10.1093/bioinformatics/bth195
- Wendel, J. F. (2000). Genome evolution in polyploids. *Plant Mol. Biol.* 42, 225–249. doi: 10.1023/A:1006392424384
- Wright, P. E., and Dyson, H. J. (2015). Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* 16, 18–29. doi: 10.1038/nrm3920
- Xie, H., Vucetic, S., Iakoucheva, L. M., Oldfield, C. J., Dunker, A. K., Uversky, V. N., et al. (2007). Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. *J. Proteome Res.* 6, 1882–1898. doi: 10.1021/pr060392u
- Yan, J., Dunker, A. K., Uversky, V. N., and Kurgan, L. (2016). Molecular recognition features (MoRFs) in three domains of life. *Mol. Biosyst.* 12, 697–710. doi: 10.1039/c5mb00640f
- Yruela, I. (2015). Plant development regulation: overview and perspectives. *J. Plant Physiol.* 182, 62–78. doi: 10.1016/j.jplph.2015.05.006
- Yruela, I., and Contreras-Moreira, B. (2012). Protein disorder in plants: a view from the chloroplast. *BMC Plant Biol.* 12:165. doi: 10.1186/1471-2229-12-165
- Yruela, I., and Contreras-Moreira, B. (2013). Genetic recombination is associated with intrinsic disorder in plant proteomes. *BMC Genomics* 14:772. doi: 10.1186/1471-2164-14-772
- Yruela, I., Oldfield, C. J., Niklas, K. J., and Dunker, A. K. (2017). Evidence for a strong correlation between transcription factor protein disorder and organismic complexity. *Genome Biol. Evol.* 9, 1248–1265. doi: 10.1093/gbe/evx073
- Yu, J., Wang, Y., Lin, W., Li, S., Li, H., Zhou, J., et al. (2005). The genomes of *Oryza sativa*: a history of duplications. *PLoS Biol.* 3:e38. doi: 10.1371/journal.pbio.0030038
- Zhao, Y., Yu, W., Hu, X., Shi, Y., Liu, Y., Zhong, Y., et al. (2018). Physiological and transcriptomic analysis revealed the involvement of crucial factors in heat stress response of *Rhododendron hainanense*. *Gene* S037, 30320–30322. doi: 10.1016/j.gene.2018.03.082
- Zheng, C., Chen, E., Albert, V. A., Lyons, E., and Sankoff, D. (2013). Ancient eudicot hexaploidy meets ancestral eusolid gene order. *BMC Genomics* 14(Suppl. 7):S3. doi: 10.1186/1471-2164-14-S7-S3
- Zhou, J., Zhao, S., and Dunker, A. K. (2018). Intrinsically disordered proteins link alternative splicing and post-translational modifications to complex cell signaling and regulation. *J. Mol. Biol.* 430, 2342–2359. doi: 10.1016/j.jmb.2018.03.028

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 © 2018 Yruela, Contreras-Moreira, Dunker and Niklas. 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.



DNA Methylation and the Evolution of Developmental Complexity in Plants

Katharina Bräutigam¹ and Quentin Cronk^{2*}

¹ Department of Biology, University of Toronto Mississauga, Mississauga, ON, Canada, ² Department of Botany, The University of British Columbia, Vancouver, BC, Canada

OPEN ACCESS

Edited by:

Annette Becker,
Justus-Liebig-Universität Gießen,
Germany

Reviewed by:

Stefan A. Rensing,
Philipps-Universität Marburg,
Germany
Daniel Schubert,
Heinrich-Heine-Universität Düsseldorf,
Germany
Claude Becker,
Gregor Mendel Institute of Molecular
Plant Biology (GMI), Austria

*Correspondence:

Quentin Cronk
quentin.cronk@ubc.ca

Specialty section:

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

Received: 18 June 2018

Accepted: 12 September 2018

Published: 04 October 2018

Citation:

Bräutigam K and Cronk Q (2018)
DNA Methylation and the Evolution
of Developmental Complexity
in Plants. *Front. Plant Sci.* 9:1447.
doi: 10.3389/fpls.2018.01447

All land plants so far examined use DNA methylation to silence transposons (TEs). DNA methylation therefore appears to have been co-opted in evolution from an original function in TE management to a developmental function (gene regulation) in both phenotypic plasticity and in normal development. The significance of DNA methylation to the evolution of developmental complexity in plants lies in its role in the management of developmental pathways. As such it is more important in fine tuning the presence, absence, and placement of organs rather than having a central role in the evolution of new organs. Nevertheless, its importance should not be underestimated as it contributes considerably to the range of phenotypic expression and complexity available to plants: the subject of the emerging field of epi-evodevo. Furthermore, changes in DNA methylation can function as a “soft” mutation that may be important in the early stages of major evolutionary novelty.

Keywords: methylome, epialleles, plant evodevo, epigenotype, epi-evodevo, histones, chromatin, Baldwin effect

INTRODUCTION – EPIGENETICS AND DEVELOPMENT

The evolution of complex organs and systems such as vasculature, rooting structures, flowers, or seeds is key to the success of land plants. Concomitantly, we observe an increase in the number of different major cell types that can range from 12 or 13 in liverworts and hornworts to 44 in flowering plants (Niklas et al., 2014). An increase in distinct cell types requires a precise and increasingly complex interpretation of the genome to initiate differentiation and maintain cell identity across mitotic divisions thus allowing for tissue and organ formation. In addition to a diversity in organ form, sessile land plants can exhibit an impressive phenotypic plasticity that contributes to their ability to colonize, grow, and reproduce in unpredictable terrestrial environments. As opposed to vertebrates with a fixed body plan, land plants are modular in construction, and individuals of the same genotype can vary in size, placement, and frequency of organs (such as leaves) depending on the environment they are exposed to. Thus, there is a complex interplay between internal and external signals. At the molecular level, land plants have a remarkable diversity of epigenetic pathways at their disposal that likely play key roles in developmental complexity, phenotypic diversity, and adaptive capacity (Bossdorf et al., 2008; Bräutigam et al., 2013; Kooke et al., 2015).

Covalent modifications of DNA and histones, together with histone variants, chromatin modulating factors, and non-coding RNAs shape the epigenetic landscape that controls development in many eukaryotes, particularly complex eukaryotes. However, it should be noted that there are general differences between animal and plant development that may affect the use of epigenetic mechanisms in gene regulation. One such difference is the relative importance of cell lineage in development. In many animals there is sustained maintenance of cell identity in

lineages, spanning many mitoses. In *Caenorhabditis elegans* for instance “rigidly determined cell lineages generate a fixed number of progeny cells of strictly specified fates” (Sulston and Horvitz, 1977). In plants, cell fate is more likely to be determined by cell-context signaling, i.e., cell position relative to its neighbors (Pennell et al., 1995). This may help to explain some plant-animal epigenetic differences. For instance, it has been suggested that gene repression by the chromatin remodeling Polycomb-group (PcG) proteins is generally less long lasting and more responsive to developmental and environmental cues in plants than it is in animals (Köhler and Grossniklaus, 2002). This in turn is perhaps related to kingdom-specific diversification of PcG proteins. An absence in plants of certain animal-specific PcG proteins that are required for long-term maintenance of gene repression in animal cell lineages is observed along with duplications of other PcG components (Köhler and Grossniklaus, 2002).

Similarly, plants differ from animals in that the plant “germline” (such as it is) is set at a late stage mostly by environmental or cell-context signaling (Whipple, 2012) whereas in animals the germline is generally determined early and is lineage based. There is, however, evidence emerging for a reduced number of stem cell divisions in axillary meristems, resulting from early set-aside, analogous to the animal germline (Burian et al., 2016). In plants the situation is complicated by the intercalation of a gametophyte generation between meiosis and gametogenesis, but in angiosperms the germline is interpreted as originating in the archesporial (meiotic fate) cells of anther or ovule (Kelliher and Walbot, 2012). The lack of separation between germ line and soma in plants may make transgenerational transmission of epialleles intrinsically more likely in plants (Jablonka and Lamb, 1998; Cronk, 2001) as the setting of the animal germline involves extensive methylation reprogramming (Lees-Murdock and Walsh, 2008). Epigenetic reprogramming in the germline of angiosperms also occurs, but it is highly specific and complex (Slotkin et al., 2009) and, unlike in animals, methylation in CG and CHG contexts is largely retained (Calarco et al., 2012).

Land plants (embryophytes) have a further major distinction from animals in that they have an alternation of haploid and diploid generations. In this, the same set of genes (in diploid vs. haploid form) generates two morphologically divergent organisms. It has been suggested that epigenetic reprogramming is likely to be involved in this extraordinary phenomenon (Cronk, 2001) and there are now studies that bear this out (Mosquna et al., 2009; Okano et al., 2009) and some that implicate DNA methylation as part of this control (Yaari et al., 2015).

In a broad evolutionary context, a particularly interesting taxonomic group are streptophyte algae (Zygnematophyceae, Coleochaetophyceae, Charophyceae, and Klebsormidiophyceae), as they comprise the lineage in which the embryophytes evolved. It is therefore of especial interest that the *Klebsormidium* and *Chara* genomes have recently been sequenced (Hori et al., 2014; Nishiyama et al., 2018), as this opens the way to future studies of methylation in these taxa. The closest algal group to the land plants is the Zygnematophyceae (Turmel et al., 2006; Timme et al., 2012; Ruhfel et al., 2014; Wickett et al., 2014; Gitzendanner et al., 2018), and further sequencing in this group will also assist

the elucidation of DNA methylation processes directly ancestral to those of the embryophytes.

The evolution of epigenetic function as it affects the evolution of plant development (“epi-evodevo”) will be a fertile field of enquiry. DNA methylation is one of the better-known epigenetic mechanisms, yet determining how the evolving methylome is linked to the evolution of plant form is still a major challenge.

CONSTITUTIVE AND FACULTATIVE EPIGENETIC CONTROL OF DEVELOPMENT

In plants, developmental control through epigenetic mechanisms can be considered either constitutive (internal signals) or facultative (external signals). Constitutive developmental control is based on internal developmental cues and is characteristic of organism-specific normal development. In contrast, facultative developmental control is based on external environmental cues and is how organisms (especially plants) respond to different environments by developmental plasticity giving rise to environment-specific phenotypes.

Epigenetic Control in Constitutive Development

Well-documented examples for constitutive epigenetic control in development relate to events central to reproduction and seed development in angiosperms/*Arabidopsis*. If disrupted, these can have detrimental effects on the formation of reproductive tissues and seeds (Chaudhury et al., 1997). Such examples include global demethylation of the genome in vegetative companion cells in the female and male gametophyte to reinforce TE silencing in the sperm and egg cells as well as the embryo (Hsieh et al., 2009; Slotkin et al., 2009; Law and Jacobsen, 2010), genomic imprinting in the nourishing tissue, the endosperm (Rodrigues and Zilberman, 2015), or extensive DNA methylation changes during seed development (Kawakatsu et al., 2017).

Double fertilization to form a triploid nutritive tissue (endosperm) is an angiosperm innovation of enormous evolutionary importance, and one that requires epigenetic controls during development. Two sperm cells are involved: one sperm nucleus fertilizes the egg to form the zygote and ultimately the embryo while the other sperm nucleus fuses with the two nuclei of the central cell to give rise to the endosperm, i.e., the tissue that nourishes the developing embryo. Intriguingly, global de-methylation is observed in the central cell of the female gametophyte prior to fertilization. This results in a marked increase of small interfering RNAs (siRNAs) likely due to the re-activation of transposons. These siRNAs are thought to migrate to the egg cell and reinforce TE silencing in the egg cell and likely the developing embryo (Hsieh et al., 2009; Law and Jacobsen, 2010). Similar demethylation and reinforcement events have also been observed in the male gametophyte (Slotkin et al., 2009).

Due to the demethylation in the central cell, maternal and paternal genomes in the endosperm differ in their DNA

methylation levels in the endosperm. This can result in the parent-of-origin dependent gene expression (genomic imprinting) which affects a number of genes including the FERTILIZATION INDEPENDENT SEEDS (FIS)-complex genes *MEDEA* and *FIS2*. Misexpression of *MEDEA*, for example, can result in seed abortion and/or abnormal embryo development (Grossniklaus et al., 1998). Theories for the emergence of genomic imprinting in plants include parental conflicts in resource allocation, co-adaptation of maternal and embryonic characteristics or dosage-dependent gene regulation, however, much remains to be learned about the biological significance and role in plant evolution (Rodrigues and Zilberman, 2015). Furthermore, following fertilization and initial embryo development, genome-wide changes in DNA methylation shape the epigenome during seed development, dormancy, and germination. During seed development, extensive hypermethylation especially in CHH context in TEs is observed which is reset during germination (Kawakatsu et al., 2017).

Another example of constitutive is provided by the role of CURLY LEAF (CLF) in normal leaf development (Goodrich et al., 1997). CLF is a polycomb group (PcG) protein, a group that functions by remodeling chromatin to maintain stable gene repression through many mitoses, i.e., marking a cell lineage. PcG proteins form modular multimeric complexes that fulfill diverse roles in development. Several genes coding for polycomb repressive complex 2 (PRC2) components have diversified in plants. The EMBRYONIC FLOWER 2-containing complex, for example, maintains vegetative development and represses reproduction while the FIS-complex prevents seed development in the absence of fertilization in *Arabidopsis* (Hennig and Derkacheva, 2009).

Epigenetic Control in Facultative Development

An example of facultative developmental control is the vernalization response. This is the plant's memory of having passed through winter, giving it the competence to flower. It involves the PcG protein VERNALIZATION 2 (VRN2) (Wood et al., 2006) along with a complex suite of epigenetic pathways including chromatin remodeling factors, histone modifications, non-coding RNAs, and DNA methylation (He, 2012). Vernalization is a facultative epigenetic response whereby environmental cues determine the developmental outcome between two phenotypic states, vegetative and flowering.

Similarly, perception of the light environment during fundamental transitions in development such as germination and photomorphogenesis (i.e., the transition to autotrophic growth in the seedling) relies on environmental signals that are translated into altered chromatin states and consequently massive transcriptional reprogramming. Light exposure perceived by photoreceptors eventually leads to reduced hypocotyl elongation, the opening of embryonic leaves (cotyledons), and chlorophyll biosynthesis. This process involves major changes in the genome organization to induce permissive chromatin states at hundreds of light-inducible genes. Such changes encompass an increase in size of the nucleus, a moderate ploidy level increase through

endoreduplication, heterochromatin condensation, and histone modifications (Bourbousse et al., 2015).

Environmental signals can be integrated into the chromatin landscape in such a way that development and genome function are fine-tuned. Morphological changes, often subtle, can be observed in postembryonic development in response to variable environmental conditions and challenges. Environmental signals can be incorporated via hormone signaling into the chromatin of vegetative meristems. This in turn can modulate root architecture or leaf development (Xiao et al., 2017). A good example of a morphological response to environmental stress is the increased leaf trichome formation in the yellow monkey flower *Mimulus guttatus* as defense against insect herbivory. Trichomes are epidermal outgrowths and a model for cell differentiation and patterning. In *M. guttatus*, trichome density is not only stimulated in herbivory exposed plants but this trait is also epigenetically transmitted to their non-stressed offspring, and is likely mediated by changes in the expression of a MYB transcription factor (Scoville et al., 2011).

A further example is the ability of *Arabidopsis* plants to fine-tune water relations epigenetically. Low relative humidity induces hypermethylation at *SPEECHLESS* (*SPCH*), a gene in the stomatal developmental pathway. This correlates with reduced *SPCH* gene expression and a reduced stomatal index (Tricker et al., 2012). Intriguingly, DNA methylation pattern at *SPCH* and stomatal phenotype were transmitted to progeny, although both were reversible under repeated stress treatment in these progeny (Tricker et al., 2013a,b).

Finally, epigenetic recombinant inbred lines (epiRILs) that are isogenic but differ in their DNA profiles are useful tools to study the effects or epigenetic variation on plant phenotype. Work in *Arabidopsis* epiRILs shows that the variation in morphological and physiological traits (e.g., flowering time or plant height) among epiRILs is comparable to that observed among natural accessions highlighting the potential of DNA methylation variation in modulating phenotype (Johannes et al., 2009; Roux et al., 2011).

Thus, DNA methylation and other modifiers of the chromatin landscape can effectively shape plant phenotype without, or prior to, genetic change; i.e., they can be considered a "conditions-sensitive ability to create diversity [...] related to both ontogenetic adaptive plasticity and evolutionary adaptation" (Jablonka, 2013).

DNA METHYLATION IN GREEN PLANTS

Methylation of cytosine (5-methylcytosine) is a common covalent modification of DNA which can be passed on across mitotic and meiotic cell divisions. While it does not alter the primary sequence of the DNA and thus the genetic information, DNA methylation plays important roles in maintenance and regulation of genome structure and function (Figure 1). For example, it contributes to the organization of chromatin into condensed heterochromatic regions, is involved in repeat silencing, and has been implicated in the regulation of gene expression and recombination. It can affect central biological processes ranging

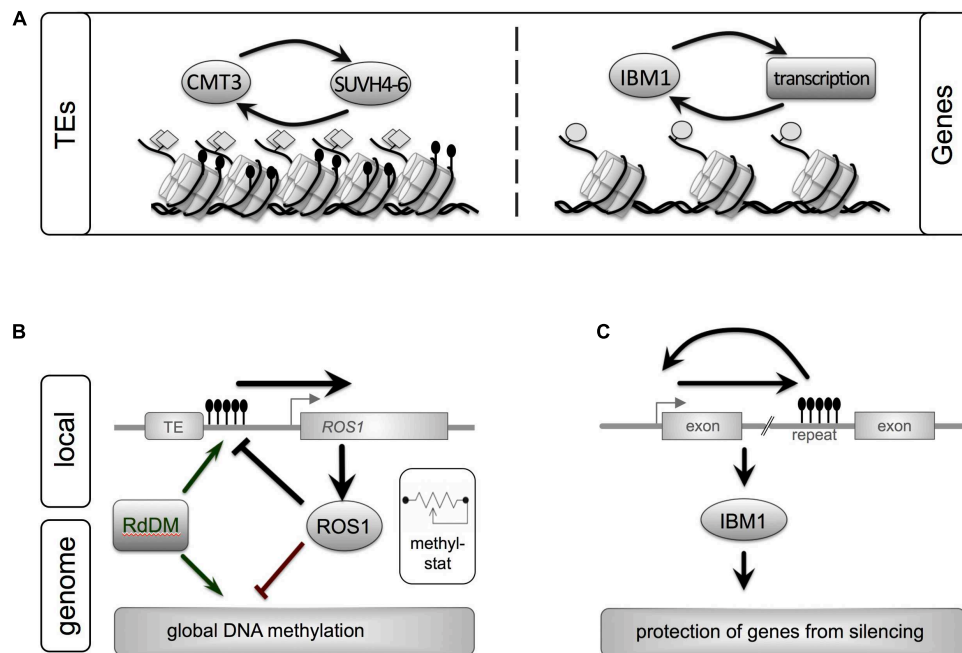


FIGURE 1 | Epigenetic maintenance of genome integrity involves both repeats and genes in plant genomes. A close interplay between genetic elements and epigenetic pathways ensures homeostasis of eu- and heterochromatin. **(A)** Silent transposons (TEs) are characterized by heterochromatic marks such as DNA methylation (black lollipops, non-CG context) and histone methylation (H3K9me2, gray diamonds). These marks are maintained by a self-reinforcing loop between CMT3, a DNA methyltransferase, and histone methyltransferases. Another feedback loop stabilizes active genes. IBM1, a histone demethylase prevents repressive H3K9me2 histone mark from spreading into genes. **(B)** ROS1, a central enzyme that removed DNA methylation marks, links local DNA methylation within its own promoter to genome-wide DNA methylation levels: acting as “methylstat.” Methylation of a sequence with the promoter of ROS1 close to a TE promotes expression of ROS. ROS1 in turn demethylates its own promoter and other regions in the genome. By monitoring the genome’s methylation status and adjusting its own expression ROS1 contributes to fine-tuning of the cell’s DNA methylation levels. Active DNA methylation is performed by RdDM or Met1. **(C)** The histone demethylase IBM1 contains a heterochromatic repeat in one of its introns. Correct IBM1 expression is crucial for the protection of genes from heterochromatic marks (H3K9me2, non-CG methylation) and thus correct genome-wide expression profiles. Reduction of DNA methylation at this repeat reduces IBM1, which leads to genome-wide hypermethylation in thousands of genes and a number of developmental defect. Gene and regulatory loops shown reflect processes in *Arabidopsis* (Lei et al., 2014; Sigman and Slotkin, 2016).

from normal cell function to genomic imprinting, regulation of development or responses to environmental cues, and is of relevance in heterosis (hybrid vigor) or polyploidization events.

In addition to 5-methylcytosine discussed herein, plant DNA can contain a number of non-canonical base modifications at low frequencies. These including various oxidized derivatives of 5-methylcytosine or *N*⁶-methyladenine (6mA) (Liu et al., 2013; Zhou et al., 2018). While such base modifications have been rarely studied in a plant evolutionary context, 6mA might emerge as an interesting epigenetic mark in plant and animal systems (Greer et al., 2015; Zhang et al., 2015; Zhou et al., 2018).

For DNA methylation, the sequence context is of relevance. Whereas animals (metazoa) are characterized predominantly by CG methylation, DNA methylation in plants occurs in all sequence contexts: symmetric CG, CHG, and asymmetric CHH (H = A, T, or C) which are set and maintained by context-specific but partially overlapping molecular pathways. A number of distinct DNA methyltransferases both generate (*de novo*), and subsequently maintain, DNA methylation at three sequence contexts: MET1 maintains CG methylation, plant-specific CHROMOMETHYLASES (CMTs) pathways target CHH (CMT2) and CHG sites (CMT3 and CMT2) in repeats and

transposons, and asymmetric CHH methylation is maintained via DRM2 through persistent *de novo* methylation (RNA-directed DNA methylation pathway, RdDM). Names here refer to proteins in *Arabidopsis* (Law and Jacobsen, 2010; Stroud et al., 2013).

DNA methylation is an ancient epigenetic mark. In the green lineage (Viridiplantae), it is found in all major taxonomic groups including Chlorophycean green algae, liverworts, mosses, ferns, gymnosperms, or angiosperms (Feng et al., 2010; Zemach et al., 2010; Takuno et al., 2016). Given its function and conservation, epigenetic regulation via DNA methylation is likely an important factor in plant evolution. The variability of methylomes among taxonomically diverse plants has recently attracted increasing attention (Niederhuth et al., 2016; Vidalis et al., 2016). However, only recently are we beginning to understand how DNA methylation patterns are shaped over evolutionary time scales, and how individual epigenetic variability contributes to phenotypic variation and adaptive potential (Bossdorf et al., 2008; Bräutigam et al., 2013; Vidalis et al., 2016). Here we bring together the current understanding of DNA methylation in plant evolution and development, drawing widely on studies from across the green plants.

DNA METHYLATION OF TRANSPOSABLE ELEMENTS

Within all plant genomes, DNA methylation shows a non-random distribution: DNA methylation is universally enriched in repetitive regions such as transposable elements (TEs), centromeric repeats, and rDNA (Slotkin and Martienssen, 2007; Feng et al., 2010; Zemach and Zilberman, 2010). Active TEs are mutagenic and can disrupt genes, regulatory regions, and affect genome integrity. Most existing TEs are, however, inactive, i.e., silenced and/or non-functional. Silencing of TEs has been proposed as one of the original functions of DNA methylation pathways (Slotkin and Martienssen, 2007; Mirouze and Vitte, 2014; Sigman and Slotkin, 2016).

Transposable elements show increased levels in DNA methylation in all sequence contexts (most prominently CG and CHG), a feature detected in almost all of the studied plant epigenomes ranging from the moss *Physcomitrella* to gymnosperms and angiosperms (Chan et al., 2008; Feng et al., 2010; Zemach et al., 2010; Niederhuth et al., 2016; Lang et al., 2018). Preferential methylation of repeats has also been reported in the distantly related green algae *Chlamydomonas*, *Chlorella*, and *Volvox* (Feng et al., 2010; Zemach and Zilberman, 2010). Work in the model plant *Arabidopsis thaliana* has shown that silent TEs adopt a distinct chromatin state, characterized by the repressive histone H3 lysine 9 dimethylation (H3K9me2) mark in combination with elevated DNA methylation levels and other histone modifications (Cokus et al., 2008; Roudier et al., 2011). This repressive heterochromatin of silent TEs is one of four major chromatin states described for the *A. thaliana* genome, and distinct from chromatin of actively transcribed genes, polycomb-repressed genes, and intergenic regions (Roudier et al., 2011).

Transposable element repression is mediated by overlapping mechanisms (double lock) including a re-inforcement loop between H3K9me2 and (non-CG) DNA methylation, and it also involves small interfering RNA (siRNA) (Slotkin and Martienssen, 2007; Roudier et al., 2011; Mirouze and Vitte, 2014; Sigman and Slotkin, 2016). While the boundaries between heterochromatic TEs and euchromatic genes are generally reinforced, heterochromatin can sometimes spread from silenced TEs and influence the expression of genes in their vicinity. Examples in *A. thaliana* include *FLOWERING WAGENINGEN* (*FWA*) or *BONSAI* (*BNS*) (Soppe et al., 2000; Saze and Kakutani, 2007). In addition, there are multiple further mechanisms by which TEs can influence the expression of genes in *cis* and in *trans* as reviewed previously (Slotkin and Martienssen, 2007; Mirouze and Vitte, 2014; Sigman and Slotkin, 2016).

FROM CONTROL OF TEs TO CONTROL OF GENES: THE METHYLATION TRANSITION

The use of DNA methylation to control TEs is nearly ubiquitous in eukaryotes and is thus apparently an ancient feature. The additional molecular inventory to use DNA methylation to

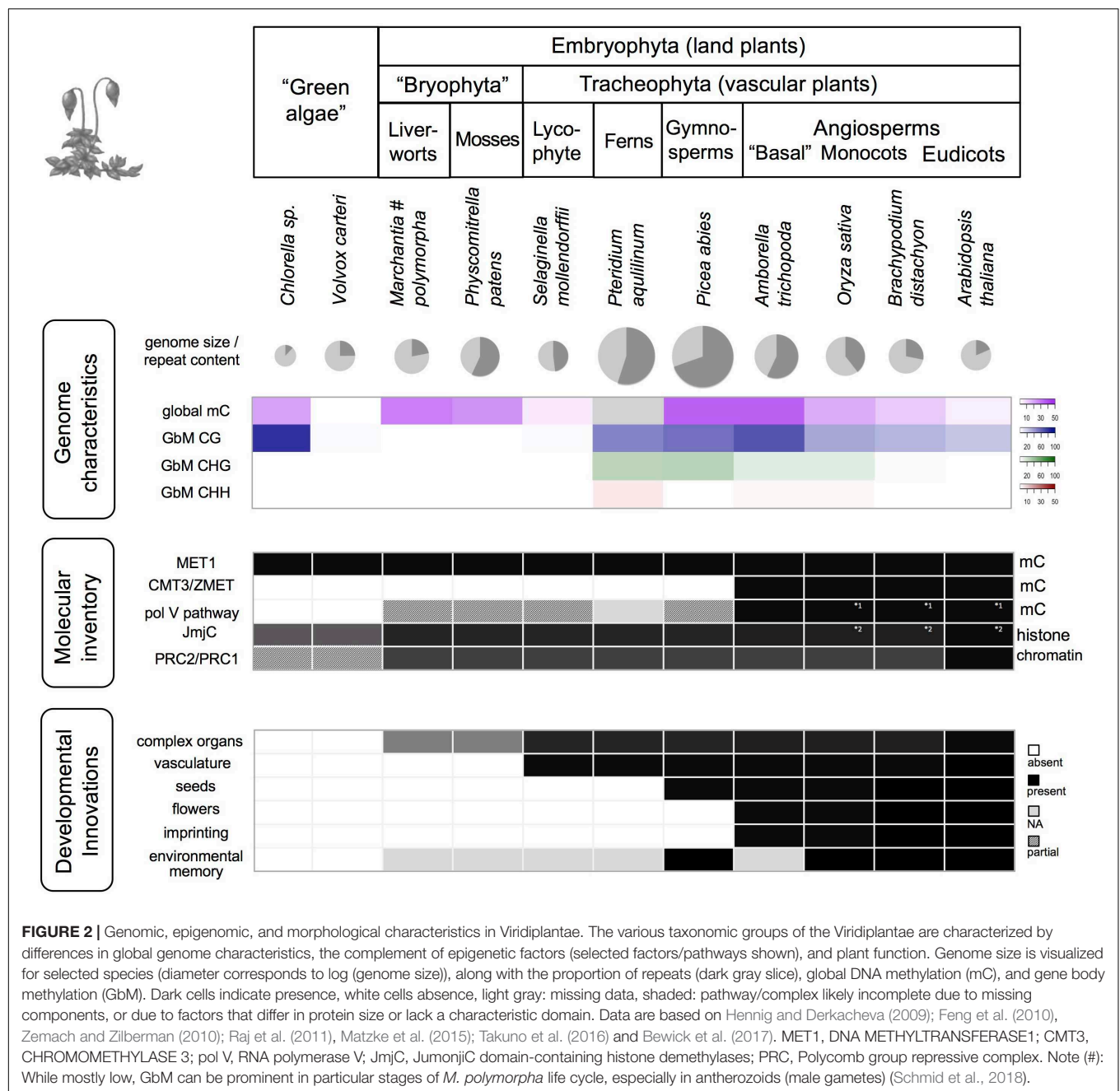
control genes, by the specific methylation of promoters and transcriptional start sites, appears to have arisen later in evolution but is apparently ubiquitous in the embryophytes (Figures 2, 3).

The moss, *Physcomitrella patens* has a DNA METHYLTRANSFERASE 1 (*MET1*) homolog as well as a CHROMOMETHYLASE (*CMT*) gene, RdDM methylases and two additional DNA methyltransferases (Malik et al., 2012). Loss of these genes results in overexpression of other genes in the moss genome implying they that they have a repressive role in gene transcriptional control.

Given the extensive level of transcriptional control by DNA methylation evident in land plants, it would be fair to ascribe considerable developmental significance to DNA methylation. It would follow that the evolution of gene expression control from TE control is one of the most significant evolutionary transitions in the emergence of complex organisms. In fact, loss of key genes, such as *MET1*, can have rather variable impacts on development. In the moss *Physcomitrella*, for instance, plants lacking the *MET1* homolog failed to produce sporophytes (Yaari et al., 2015). This drastic effect has been ascribed to impaired gamete development, fertilization or early steps in embryo development due to concomitant loss of CG methylation. In contrast, loss of the *MET1* homolog, had, surprisingly, no effect on gametophyte development (Yaari et al., 2015); however, treatment of the gametophyte with the methyltransferase inhibitor zebularine does produce abnormal phenotypes (Malik et al., 2012). These observations in *Physcomitrella* indicate strong life-cycle specificity of DNA methylation effects as well as partial redundancy in the DNA methylation machinery (Malik et al., 2012; Yaari et al., 2015).

Similarly, *Arabidopsis* mutants with altered DNA methylation levels show diverse phenotypic characteristics. Lack of functional *AtMET1* which results in strongly reduced CG methylation, leads to a number of abnormalities including small plant size, reduced fertility, changes in flowering time, or altered floral morphologies (Finnegan et al., 1996; Kinkel et al., 2003; Mathieu et al., 2007). Here, the late flowering phenotype can be attributed to ectopic *FWA* expression due to DNA methylation loss at a TE upstream of the gene (Soppe et al., 2000). Moreover, loss of *MET1* in *Arabidopsis* results in impaired development in a significant proportion of embryos (Xiao et al., 2006). These embryos show misregulation of gene expression, abnormal patterns of cell division (planes and number or cell divisions) or improperly formed auxin gradients (Xiao et al., 2006). While some *Arabidopsis* mutants with lesions in individual components of the DNA methylation machinery such as the DNA methyltransferases *CMT3*, *DRM1*, or *DRM2* grow like wildtype plants, simultaneous loss of multiple components results in developmental defects (e.g., *drm1 drm2 cmt3* triple mutant with loss of non-CG methylation (Cao and Jacobsen, 2002; Chan et al., 2006), or enhances phenotypic abnormalities: *met1 cmt3* double mutant (Xiao et al., 2006).

The mild defects observed in several individual *Arabidopsis* mutants thus likely reflect redundancies of DNA methylation pathways and the generally low DNA methylation levels in



this plant (Figure 2). Plant species with larger genomes, higher repeat content, and global DNA methylation levels such as rice or maize show more severe phenotypes or lethality in DNA methylation mutants (Erhard et al., 2009; Hu et al., 2014). For example, lack of OsMET1 in rice resulted in abnormal seeds and seedling lethality (Hu et al., 2014), and in contrast to *Arabidopsis*, rice chromomethylase (*cmt3a*) mutants produced less biomass, showed low fertility and were characterized by complex expression changes of cellular genes and mobilization of TEs (Cheng et al., 2015). These examples show that while DNA methylation retains a central function in TE management, it also plays key roles in the regulation of developmentally important

genes, likely as part of a complex regulatory network, and with built in redundancies.

DIVERSIFICATION OF DNA METHYLATION PATHWAYS

While DNA methyltransferase 1 (MET1), CMT, and RdDM functionality is present in all land plants (i.e., setup and maintenance of DNA methylation in all sequence contexts), there is evident evolutionary diversification in several DNA methylation pathways, indicating an expansion of their roles. For

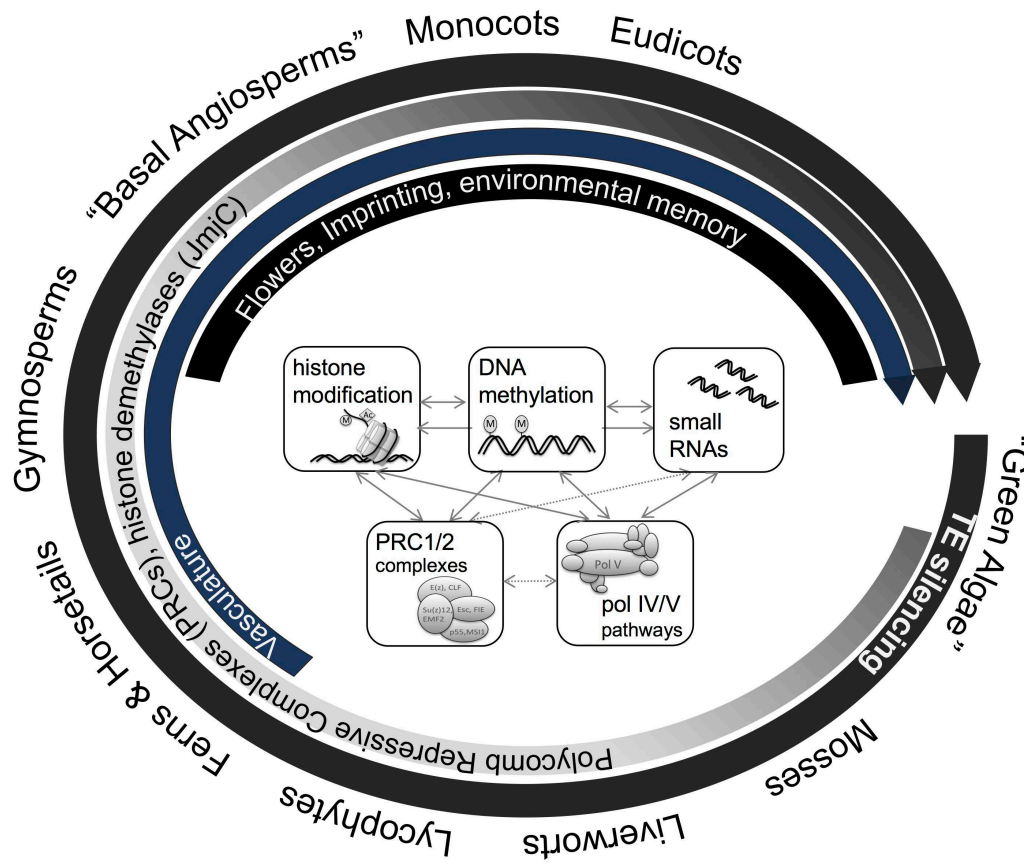


FIGURE 3 | Macro-epi-evodevo. The evolution of land plants is characterized by various developmental innovations, which are accompanied by emergence, co-option, and divergence of various genome- and epigenome-related processes. Plants uniformly employ DNA methylation and related pathways for Transposon (TE) silencing, genomes of ferns and more evolved plants show gene body methylation, and histone modifying enzymes (JmjC histone demethylases) as well as chromatin modifying complexes (PRC1/2) have diversified during the evolution of land plants. More recently, new pathways such as CMT3 or pol V pathways have emerged during the evolution of angiosperms which are involved in DNA methylation in non-symmetric sequence context or have been hypothesized to play a role in genome-reduction after polyploidization. In parallel, new phenomena such as genomic imprinting, complex organ development, and environmental memory occur in angiosperms.

example, the complex pol V branch of RdDM pathway (**Figure 2**) is only fully functional in angiosperms and has been hypothesized to play a role in diploidization and genome reduction after whole-genome duplication shock (Matzke et al., 2015). Similarly, CMT3, involved in the maintenance of non-CG methylation through a reinforcement loop with histone methylation (H3K9me2) is angiosperm-specific and can be counteracted by an angiosperm-specific histone demethylase (Increase in *BONSAI* Methylation 1, IBM1) that prevents spreading of DNA methylation. More generally, histone demethylases and PRC components have diversified gradually during land plant evolution (shown for JmjC type in **Figure 2**) potentially contributing to the regulation of increased developmental complexity and extensive interactions with the environment. **Table 1** gives some examples of the control of developmental pathways by targeted methylation or demethylation.

Further complexity in DNA methylation pathways result from linking a repeat sequence to the plant's methylation status. An interesting example of this, for the control and fine-tuning of

gene expression and development in *Arabidopsis*, is provided by *ROS1* (*REPRESSOR OF SILENCING1*) and *IBM1* (**Figures 1B,C**). Both can act as epigenetic rheostat or “methylstat” to establish a genome-specific DNA methylation equilibrium. *ROS1* encodes a DNA demethylase, an enzyme that removes DNA methylation catalytically. The *ROS1* enzyme functions genome-wide and counteracts DNA methylation pathways such as RdDM. In the *ROS1* promoter, MEMS (DNA methylation monitoring sequence), a sequence adjacent to a *Helitron* TE, is critical for regulating the expression of *ROS1*. Somewhat counter-intuitively, *ROS1* expression is promoted when this sequence is methylated (e.g., by RdDM) and inhibited by demethylation (e.g., by *ROS1* itself). Upon expression, the *ROS1* enzyme demethylates its own promoter thus reducing its own expression. Reduced *ROS1* activity allows then for increased *ROS1* promoter methylation and expression until an equilibrium is reached (Sigman and Slotkin, 2016).

Increase in *BONSAI* Methylation 1 encodes a histone demethylase which is involved in preventing spread of DNA

TABLE 1 | Some examples of developmental genes regulated by, or affected by, methylation status.

Gene	Developmental pathway	Notes	Reference
<i>AGAMOUS (AG)/SUPERMAN (SUP)</i>	Floral whorls	Mutants with hypermethylation in <i>AG</i> and <i>SUP</i> gene bodies display floral abnormalities	Jacobsen et al., 2000
<i>APOLO</i> lncRNA	Auxin	Auxin triggers demethylation of <i>APOLO</i> DNA and lncRNA production	Ariel et al., 2014
<i>CYCLOIDEA (CYC)</i>	Floral zygomorphy	Naturally occurring hypermethylated teratomorph of <i>Linaria</i> abolishes <i>CYC</i> expression and results in change of floral symmetry	Cubas et al., 1999
<i>EPIDERMAL PATTERNING FACTOR 2 (EPF2)</i>	Stomatal development	Normal development requires demethylation by <i>ROS1</i>	Yamamuro et al., 2014
<i>SPEECHLESS (SPCH)</i>	Stomatal patterning (stomatal index)	Low relative humidity induces methylation at <i>SPCH</i>	Tricker et al., 2012
<i>FLOWERING WAGENINGEN (FWA)</i>	Flowering time	Epialleles of <i>FWA</i> confer late flowering phenotype	Soppe et al., 2000; Kinoshita et al., 2004
<i>MEDEA (MEA)</i>	Seed development	Allele-specific demethylation by <i>DEMETER (DME)</i> required for the establishment of self-imprinting by <i>MEDEA</i>	Gehring et al., 2006
<i>PHABULOSA (PHB)/PHAVOLUTA (PHV)</i>	Leaf development; polarity	<i>PHB</i> and <i>PHV</i> coding sequences have functionally important methylation downstream of microRNA binding site	Bao et al., 2004
<i>RIPENING INHIBITOR (RIN)/NON-RIPENING (NOR)</i>	Fruit ripening in tomato	Demethylation of ripening genes, including <i>RIN</i> and <i>NOR</i> , by the tomato homolog of <i>ROS1</i> , is required for normal ripening	Lang et al., 2017

All genes are from *Arabidopsis* except *CYC* (*Linaria*) and *RIN/NOR* (tomato).

methylation into genes and regulating genome-wide DNA methylation patterns via a feedback loop. *IBM1* contains a heterochromatic repeat in one of its introns. Reduced DNA methylation of this repeat element (e.g., in a mutant background) results in reduced *IBM1* expression (improper polyadenylation), followed by increased DNA methylation (mCHG) (Lei et al., 2014; Sigman and Slotkin, 2016; Zhang et al., 2018).

While the exact mechanisms described here for examples in *Arabidopsis* are likely species-specific, similar mechanisms may exist in other plants. The expansion of epigenetic studies from *Arabidopsis* to other systems will be essential to understand which mechanisms are evolutionarily conserved and which are species specific.

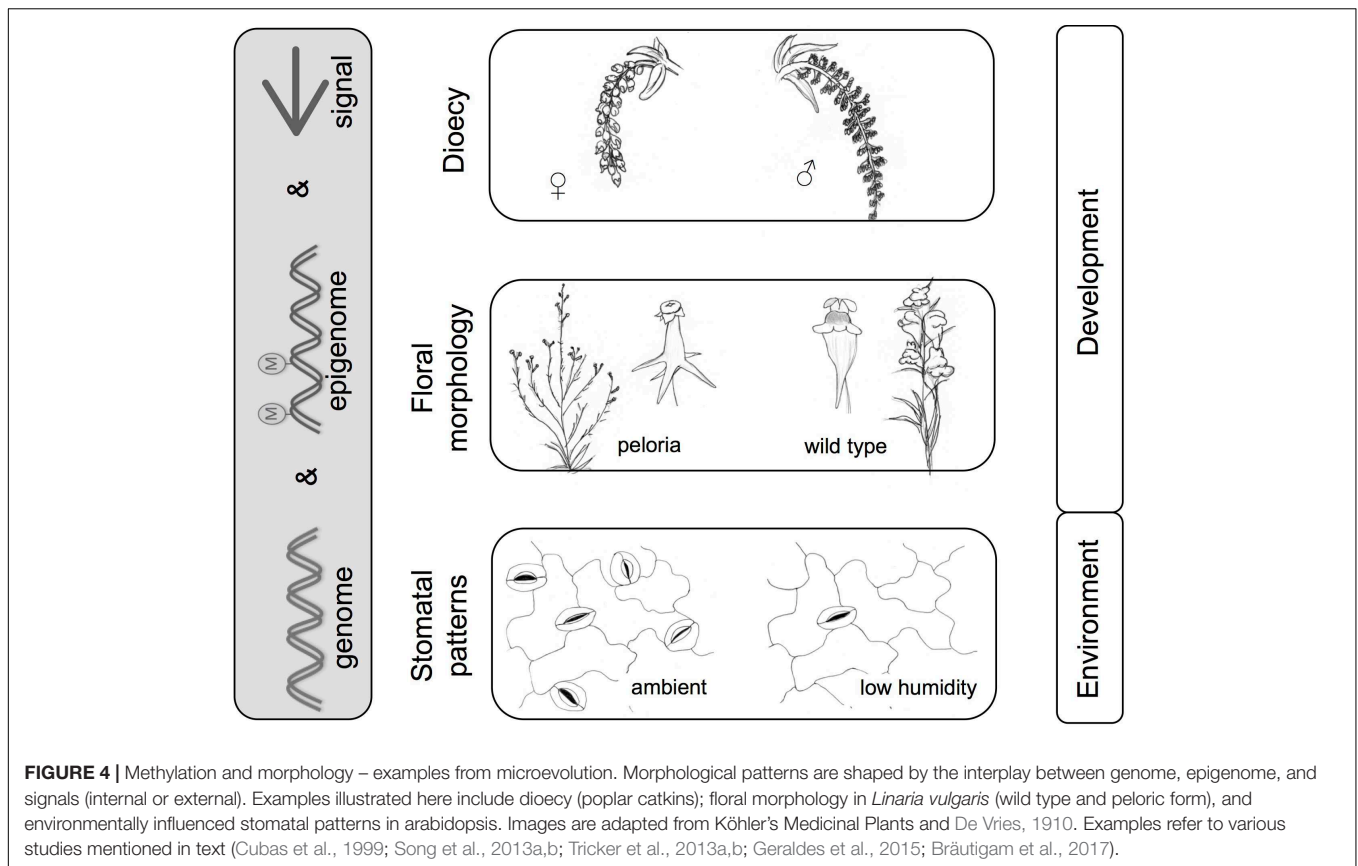
DNA METHYLATION OF THE GENE BODY

Methylation of actively transcribed genes (predominantly exons: gene body methylation, GbM) is another feature occurring in plant genomes. Although DNA methylation at promoters and transcriptional start sites (gene promoter methylation) has been associated with transcriptional repression, GbM does not generally repress gene expression. Instead, GbM genes are

typically expressed constitutively in a wide range of tissues and conditions at moderate levels (housekeeping genes) (Zemach et al., 2010; Bewick et al., 2017). Body-methylated genes thus represent a distinct set of genes, and comprise, for example, approximately 18% of the genes in *Arabidopsis thaliana* ecotype Columbia (Col-0) (Takuno and Gaut, 2012).

Whereas methylation of TEs, and methylation of gene control regions, are apparently ancient in plants, it appears that GbM has expanded in plants more recently, as GbM appears to be minimal in early diverging lineages of land plants (Takuno et al., 2016). However, recent work (Schmid et al., 2018) examining various stages of the *Marchantia* life cycle has shown that GbM is not absent from *Marchantia*, merely prominent in particular stages (it is abundant in the antherozoids). A recent study of *Physcomitrella patens* also revealed that ca. 5.7% protein-coding genes have at least one methylated position in their gene body (Lang et al., 2018). In animals, the situation may be rather different: gene body methylation appears to be general and ancient (Dixon et al., 2016).

Despite ongoing work, the function of GbM still remains mysterious. DNA methylation is potentially mutagenic as spontaneous deamination can convert 5-methylcytosine into thymine, thus the retention of GbM likely comes at a cost. Nevertheless, GbM genes share conserved features and their



occurrence spans at least 400 Myr of land plant evolution (Zilberman, 2017). Given the conservation of GbM in evolution (Takuno and Gaut, 2013), GbM genes might be expected to play important roles in plant development. This is supported by *Arabidopsis* mutants. Mutants with severely reduced GbM but largely intact TE methylation show a number of morphological and developmental defects, a pattern that is even observed over progressive generations (Mathieu et al., 2007; Stroud et al., 2013). However, secondary loss of GbM in two Brassicaceae species, *C. planisiliqua* and *E. salsugineum*, indicates that GbM is non-essential over evolutionary time (Bewick et al., 2016).

This paradox, that GbM is likely important but also dispensable, remains to be resolved. Numerous potential functions for GbM have been proposed. These include: (1) involvement in accurate transcription and splicing; (2) the repression of cryptic intragenic promoters (Jeltsch, 2010; Zemach and Zilberman, 2010; Takuno and Gaut, 2012); and (3) sheltering genes from TE insertions (To et al., 2011) while functionless alternatives have also been discussed (Roudier et al., 2009).

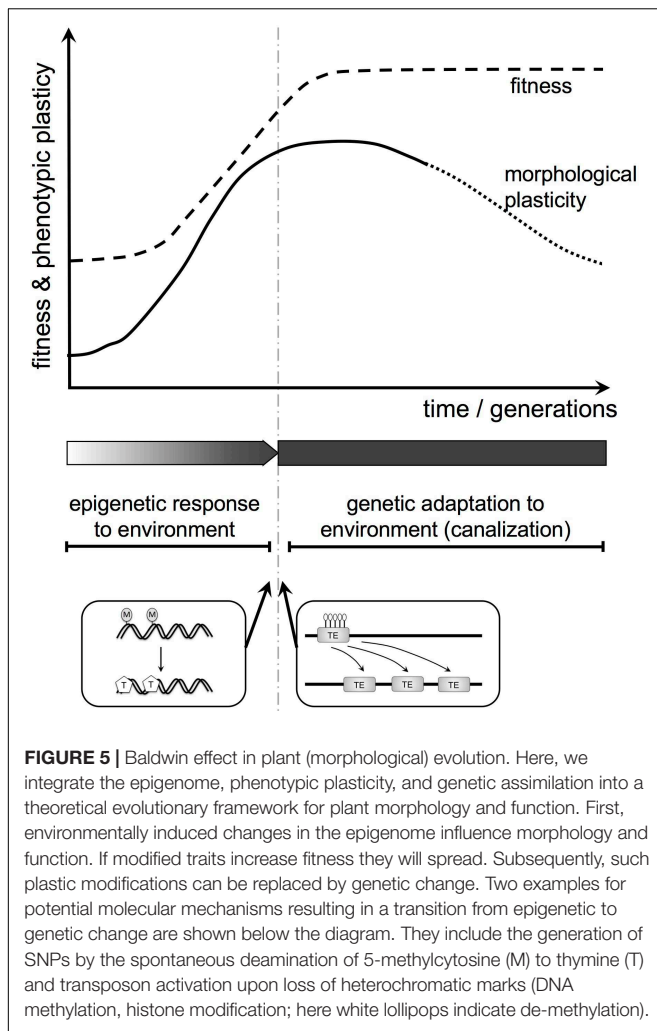
METHYLATION AND THE EVOLUTION OF MORPHOLOGY: EXAMPLES

Dioecy

One of the best supported roles for the action of DNA methylation in plant evolution is provided by the evolution of

separate sexes in plants from a cosexual ancestor (Figure 4). Dioecy has evolved independently in multiple lineages and it has been suggested that methylation might be a key mechanism (Gorelick, 2003). One genus in which this idea has received support is *Populus*. In the Chinese white poplar (*Populus tomentosa*) sex-specific methylation has been implicated (Song et al., 2013a). More recently, a genomic characterization of the *Populus trichocarpa* sex locus (Geraldès et al., 2015) found a methyltransferase (poplar *MET1* homolog) present at the sex determining region (SDR). It is also of interest that a possible methyltransferase has been noted at the SDR of strawberry (Tennessen et al., 2016), an observation that would merit further investigation. Further support for the involvement of methylation in sex determination in poplar has come from the finding that another gene at the poplar SDR, the poplar homolog of the *Arabidopsis* Response Regulator 17 (*ARR17*), is markedly sex-specifically methylated (Bräutigam et al., 2017). Male individuals have generally stronger methylation, including at the putative promoter region.

Evidence is also accruing in other systems. Thus, in the dioecious *Silene latifolia* (white campion), treatments with demethylating agents can alter sex expression in the flowers, converting male flowers to hermaphrodite ones (Janoušek et al., 1996). Other dioecy systems that implicate methylation include persimmons (Akagi et al., 2016; Henry et al., 2018) and papaya (Zhang et al., 2008; Liu et al., 2018). In the latter, CHH-context methylation of *HUA1*, an *AGAMOUS* (*AG*)



regulator, is associated with sex reversal in papaya. In the Cucurbitaceae, a family known for its flexible sexual systems (from hermaphroditism and monoecy to dioecy), methylation is also implicated in floral sex determination (Martin et al., 2009; Lai et al., 2017). There is now no doubt that sex-specific methylation is a mechanism that has been employed independently multiple times in the evolution of dioecy.

Peloria

Linaria vulgaris (toadflax) has a remarkable floral mutant, first characterized by Linnaeus, called peloria. This results from the ventralization of flowers, leading to flowers with 5 (ventral) spurs instead of one. It is caused by abolition of function of the dorsal identity gene *CYCLOIDEA*. In now classic work this was shown to result from *CYCLOIDEA* gene repression by methylation (Cubas et al., 1999; Figure 4). Teratomorphs derived in this way can persist by vegetative reproduction (clump formation by root buds), but produce little seed so the mutant is semi-lethal as regards sexual reproduction. The inheritance of this feature was investigated by De Vries in “Die Mutationstheorie” (De Vries, 1910). De Vries divided peloric individuals into two types: (1)

hemipelagic in which a mixture of peloric and wild type flowers occurred in an inflorescence, sometimes with intermediates, and (2) fully peloric, in which all flowers in the inflorescence are peloric. As hemipelagic plants have some wild type flowers, they are fertile and this likely explains the wide persistence of the potential for abnormal methylation within the species. The epiallele is heritable, although largely recessive. Fully peloric plants crossed with wild-type produce mostly wild-type with a low frequency of hemipelorics (Cubas et al., 1999).

Linaria vulgaris therefore seems (as De Vries puts it) to have “an inherited semi-latent character, which manifests itself from time to time” (De Vries, 1910). Given that similar phenotypes are also present in *Linaria purpurea* (Rudall and Bateman, 2003) it may be that the latent character is phylogenetically conserved in the genus and thus a “latent homology” (Cronk, 2002). Currently unexplained are the pleiotropic effects of *CYC* methylation on other aspects of morphology besides floral ventralization. Fully peloric plants often have a strongly branched inflorescence as opposed to the simple or near simple raceme of the wild type. There are also reported abnormalities of the pollen and capsule (De Vries, 1910). Even though peloria of *Linaria* is a well-studied epimutation, it is far from giving up all its secrets.

Evolutionarily, this epimutation might seem to be a dead-end. However, its persistence, perhaps through the transmission of weak epialleles via hemipelagic forms, potentially allows further mutations (for instance in corolla shape) to be occasionally expressed in a peloric or hemipelagic background. It is not hard to conceive that this could lead to an alternative pollination niche, reduction in infertility and eventually to speciation. In such a case, the “soft” mutation provided by methylation will have been crucial. Eventually the epigenetic basis could be replaced by genetic loss of function mutation, in which case the initial involvement of methylation in the evolution of a new species will be hard if not impossible to discern. A genetic loss of function mutation is less promising as a starting point, as it is likely to be lethal and to be quickly purged from the population. It is worth noting that new lineages have indeed formed from peloria-type changes. An example of this is *Cadia* (Citerne et al., 2006) a peloric legume (although here the peloria is due to dorsalization rather than ventralization and there is no evidence of methylation being involved).

THE BALDWIN EFFECT – A THEORETICAL MODEL

There is a possible role for environmentally regulated epigenetic control (including methylation) in plant evolution through the Baldwin effect (in the broad sense, including genetic assimilation (Pigliucci et al., 2006; Figure 5). The Baldwin effect is usually associated with animals: behavior, being plastic, can change as a learning response to environmental cues, allowing colonization of, and adaptation to, a new environment. Any mutation that gives a genetic predisposition to the changed behavior may be favored as it reduces the cost of learning and increases adaptation to the new environment. Thus, learned behavior can become instinctive behavior.

The same mechanism can apply in plants but here the plastic response is morphological not behavioral. Many complex animals (for instance most vertebrates) show remarkably little morphological phenotypic plasticity. Their plasticity tends to reside in the “extended phenotype,” for instance in behavior. In contrast, the same plant genotype in different environments can look dramatically different, differences that may be stabilized epigenetically. Thus, the Baldwin effect, while a driver of behavioral evolution in animals, may be a driver of morphological evolution in plants. It is a potential mechanism for replacing a phenotypic and epigenetic response to environment with a genetic adaptation to environment (Figure 5). These considerations are currently speculative but provide a conceptual framework for future experimental work.

CONCLUDING REMARKS – THE METHYLATION TOOLBOX AND ITS APPLICATION

If, as seems likely, the regulation of genes by methylation evolved from the universal eukaryotic feature of TE defense, then this is an evolutionary change with considerable implications for the evolution of embryophytes. The transition of DNA methylation function from regulation of TEs to regulation of genes, is one of the great evolutionary transitions in the evolution of complex plant life on earth. Methylation of DNA now supplies an important toolbox for fine tuning development, especially when considering its interaction with other epigenetic mechanisms: histone methylation and small RNAs.

However, because of the inherent lability and reversibility of methylation, it may be an “evolutionary sandbox” for the soft exploration of developmental space (Cronk, 2001). It allows for added phenotypic plasticity and infraspecific diversity in the expression of plant morphology (Bossdorf et al., 2008; Bräutigam et al., 2013). Later in evolution, methylation-controlled traits could become hardwired through direct sequence changes, aided by the fact that methylated DNA mutates at a higher rate.

Hypermethylation-based epimutations of a gene, as in peloric *Linaria*, could function as gene knock-downs to allow adaptation prior to gene knock-out and loss. If this is true, then it may be hard to assess the importance of methylation in evolution,

as methylation might have been involved in the early stages of the evolution of a number of important traits, but might not be evident now.

Another evolutionarily significant difference between mutation and epimutation lies in exposure to selection (Cronk, 2001; van der Graaf et al., 2015; Vidalis et al., 2016). A conventional loss-of-function recessive mutation will tend to be very rare in a population and in an outbreeding population will be unlikely to occur as the double recessive necessary to generate a selectable phenotype. An epimutation, however, may possibly affect both alleles simultaneously and thus be immediately exposed to selection. There are many elegant studies of the effect of selection on naturally occurring gene mutations, but, with some notable exceptions, e.g., in rice (Zheng et al., 2017), epimutations have been comparatively neglected. This is a challenge for the future.

AUTHOR CONTRIBUTIONS

Both authors wrote the paper, made a substantial intellectual contribution to the work, and approved it for publication.

FUNDING

Research in the laboratory of QC was funded by a Discovery Grant (RGPIN-2014-05820) from the Natural Sciences and Engineering Research Council of Canada (NSERC). Research in the laboratory of KB was funded by a Discovery Grant (RGPIN-2017-06552) from NSERC, the Canada Foundation for Innovation, and the University of Toronto.

ACKNOWLEDGMENTS

We would like to thank Verónica S. Di Stilio, Annette Becker, and Natalia Pabón-Mora, editors of this collection (Genetic Regulatory Mechanisms Underlying Developmental Shifts in Plant Evolution), for their invitation to contribute. We also apologize to those authors of notable papers on methylation in plant evolution for which space constraints did not allow inclusion.

REFERENCES

- Akagi, T., Henry, I. M., Kawai, T., Comai, L., and Tao, R. (2016). Epigenetic regulation of the sex determination gene MeGI in polyploid persimmon. *Plant Cell* 28, 2905–2915. doi: 10.1105/tpc.16.00532
- Ariel, F., Jegu, T., Latrasse, D., Romero-Barrios, N., Christ, A., Benhamed, M., et al. (2014). Noncoding transcription by alternative RNA polymerases dynamically regulates an auxin-driven chromatin loop. *Mol. Cell* 55, 383–396. doi: 10.1016/j.molcel.2014.06.011
- Bao, N., Lye, K. W., and Barton, M. K. (2004). MicroRNA binding sites in *Arabidopsis* class III HD-ZIP mRNAs are required for methylation of the template chromosome. *Dev. Cell* 7, 653–662. doi: 10.1016/j.devcel.2004.10.003
- Bewick, A. J., Ji, L., Niederhuth, C. E., Willing, E. M., Hofmeister, B. T., Shi, X., et al. (2016). On the origin and evolutionary consequences of gene body DNA methylation. *Proc. Natl. Acad. Sci. U.S.A.* 113, 9111–9116. doi: 10.1073/pnas.1604666113
- Bewick, A. J., Niederhuth, C. E., Ji, L., Rohr, N. A., Griffin, P. T., Leebens-Mack, J., et al. (2017). The evolution of CHROMOMETHYLASES and gene body DNA methylation in plants. *Genome Biol.* 18:65. doi: 10.1186/s13059-017-1195-1
- Bossdorf, O., Richards, C. L., and Pigliucci, M. (2008). Epigenetics for ecologists. *Ecol. Lett.* 11, 106–115. doi: 10.1111/j.1461-0248.2007.01130.x
- Bourbousse, C., Mestiri, I., Zabulon, G., Bourge, M., Formigini, F., Koini, M. A., et al. (2015). Light signaling controls nuclear architecture reorganization during seedling establishment. *Proc. Natl. Acad. Sci. U.S.A.* 112, E2836–E2844. doi: 10.1073/pnas.1503512112
- Bräutigam, K., Soolanayakanahally, R., Champigny, M., Mansfield, S., Douglas, C., Campbell, M. M., et al. (2017). Sexual epigenetics: gender-specific methylation

- of a gene in the sex determining region of *Populus balsamifera*. *Sci. Rep.* 7:45388. doi: 10.1038/srep45388
- Bräutigam, K., Vining, K. J., Lafon-Placette, C., Fossdal, C. G., Mirouze, M., Marcos, J. G., et al. (2013). Epigenetic regulation of adaptive responses of forest tree species to the environment. *Ecol. Evol.* 3, 399–415. doi: 10.1002/ece3.461
- Burian, A., de Reuille, P. B., and Kuhlemeier, C. (2016). Patterns of stem cell divisions contribute to plant longevity. *Curr. Biol.* 26, 1385–1394. doi: 10.1016/j.cub.2016.03.067
- Calarco, J. P., Borges, F., Donoghue, M. T. A., Van Ex, F., Jullien, P. E., Lopes, T., et al. (2012). Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. *Cell* 151, 194–205. doi: 10.1016/j.cell.2012.09.001
- Cao, X., and Jacobsen, S. E. (2002). Role of the *Arabidopsis* DRM methyltransferases in de novo DNA methylation and gene silencing. *Curr. Biol.* 12, 1138–1144. doi: 10.1016/S0960-9822(02)00925-9
- Chan, A. P., Melake-Berhan, A., O'Brien, K., Buckley, S., Quan, H., Chen, D., et al. (2008). The highest-copy repeats are methylated in the small genome of the early divergent vascular plant *Selaginella moellendorffii*. *BMC Genomics* 9:282. doi: 10.1186/1471-2164-9-282
- Chan, S. W., Henderson, I. R., Zhang, X., Shah, G., Chien, J. S., and Jacobsen, S. E. (2006). RNAi, DRD1, and histone methylation actively target developmentally important non-CG DNA methylation in *Arabidopsis*. *PLoS Genet.* 2:e83. doi: 10.1371/journal.pgen.0020083
- Chaudhury, A. M., Ming, L., Miller, C., Craig, S., Dennis, E. S., and Peacock, W. J. (1997). Fertilization-independent seed development in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 94, 4223–4228. doi: 10.1073/pnas.94.8.4223
- Cheng, C., Tarutani, Y., Miyao, A., Ito, T., Yamazaki, M., Sakai, H., et al. (2015). Loss of function mutations in the rice chromomethylase OsCMT3a cause a burst of transposition. *Plant J.* 83, 1069–1081. doi: 10.1111/tpj.12952
- Citerne, H. L., Pennington, R. T., and Cronk, Q. C. B. (2006). An apparent reversal in floral symmetry in the legume *Cadia* is a homeotic transformation. *Proc. Natl. Acad. Sci. U.S.A.* 103, 12017–12020. doi: 10.1073/pnas.0600986103
- Cokus, S. J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C. D., et al. (2008). Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* 452, 215–219. doi: 10.1038/nature06745
- Cronk, Q. C. B. (2001). Plant evolution and development in a post-genomic context. *Nat. Rev. Genet.* 2, 607–619. doi: 10.1038/35084556
- Cronk, Q. C. B. (2002). "Perspectives and paradigms in plant evo-devo," in *Developmental Genetics and Plant Evolution*, eds Q. C. B. Cronk, R. M. Bateman, and J. A. Hawkins (Boca Raton, FL: CRC Press).
- Cubas, P., Vincent, C., and Coen, E. (1999). An epigenetic mutation responsible for natural variation in floral symmetry. *Nature* 401, 157–161. doi: 10.1038/43657
- De Vries, H. (1910). *The Mutation theory: Experiments and Observations on the Origin of Species in the Vegetable Kingdom*, Vol. 2. Chicago, IL: Open Court Publishing Company.
- Dixon, G. B., Bay, L. K., and Matz, M. V. (2016). Evolutionary consequences of DNA methylation in a basal metazoan. *Mol. Biol. Evol.* 33, 2285–2293. doi: 10.1093/molbev/msw100
- Erhard, K. F. Jr., Stonaker, J. L., Parkinson, S. E., Lim, J. P., Hale, C. J., Hollick, J. B., et al. (2009). RNA polymerase IV functions in paramutation in *Zea mays*. *Science* 323, 1201–1205. doi: 10.1126/science.1164508
- Feng, S., Cokus, S. J., Zhang, X., Chen, P. Y., Bostick, M., Goll, M. G., et al. (2010). Conservation and divergence of methylation patterning in plants and animals. *Proc. Natl. Acad. Sci. U.S.A.* 107, 8689–8694. doi: 10.1073/pnas.1002720107
- Finnegan, E. J., Peacock, W. J., and Dennis, E. S. (1996). Reduced DNA methylation in *Arabidopsis thaliana* results in abnormal plant development. *Proc. Natl. Acad. Sci. U.S.A.* 93, 8449–8454. doi: 10.1073/pnas.93.16.8449
- Gehring, M., Huh, J. H., Hsieh, T. F., Penterman, J., Choi, Y., Harada, J. J., et al. (2006). DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. *Cell* 124, 495–506. doi: 10.1016/j.cell.2005.12.034
- Geraldes, A., Hefer, C. A., Capron, A., Kolosova, N., Martinez-Nunez, F., Soalanayakanahally, R. Y., et al. (2015). Recent Y chromosome divergence despite ancient origin of dioecy in poplars (*Populus*). *Mol. Ecol.* 24, 3243–3256. doi: 10.1111/mec.13126
- Gitzendanner, M. A., Soltis, P. S., Wong, G. K., Ruhfel, B. R., and Soltis, D. E. (2018). Plastid phylogenomic analysis of green plants: a billion years of evolutionary history. *Am. J. Bot.* 105, 291–301. doi: 10.1002/ajb2.1048
- Goodrich, J., Puangsomlee, P., Martin, M., Long, D., Meyerowitz, E. M., and Coupland, G. (1997). A polycomb-group gene regulates homeotic gene expression in *Arabidopsis*. *Nature* 386, 44–51. doi: 10.1038/386044a0
- Gorelick, R. (2003). Evolution of dioecy and sex chromosomes via methylation driving Muller's ratchet. *Biol. J. Linn. Soc.* 80, 353–368. doi: 10.1046/j.1095-8312.2003.00244.x
- Greer, E. L., Blanco, M. A., Gu, L., Sendinc, E., Liu, J., Aristizábal-Corralles, D., et al. (2015). DNA methylation on N6-adenine in *C. elegans*. *Cell* 161, 868–878. doi: 10.1016/j.cell.2015.04.005
- Grossniklaus, U., Vielle-Calzada, J. P., Hoepfner, M. A., and Gagliano, W. B. (1998). Maternal control of embryogenesis by MEDEA, a polycomb group gene in *Arabidopsis*. *Science* 280, 446–450. doi: 10.1126/science.280.5362.446
- He, Y. (2012). Chromatin regulation of flowering. *Trends Plant Sci.* 17, 556–562. doi: 10.1016/j.tplants.2012.05.001
- Hennig, L., and Derkacheva, M. (2009). Diversity of Polycomb group complexes in plants: same rules, different players? *Trends Genet.* 25, 414–423. doi: 10.1016/j.tig.2009.07.002
- Henry, I. M., Akagi, T., Tao, R., and Comai, L. (2018). One hundred ways to invent the sexes: theoretical and observed paths to dioecy in plants. *Annu. Rev. Plant Biol.* 69, 553–575. doi: 10.1146/annurev-arplant-042817-040615
- Hori, K., Maruyama, F., Fujisawa, T., Togashi, T., Yamamoto, N., Seo, M., et al. (2014). *Klebsormidium flaccidum* genome reveals primary factors for plant terrestrial adaptation. *Nat. Commun.* 5:3978. doi: 10.1038/ncomms4978
- Hsieh, T. F., Ibarra, C. A., Silva, P., Zemach, A., Eshed-Williams, L., Fischer, R. L., et al. (2009). Genome-wide demethylation of *Arabidopsis* endosperm. *Science* 324, 1451–1454. doi: 10.1126/science.1172417
- Hu, L., Li, N., Xu, C., Zhong, S., Lin, X., Yang, J., et al. (2014). Mutation of a major CG methylase in rice causes genome-wide hypomethylation, dysregulated genome expression, and seedling lethality. *Proc. Natl. Acad. Sci. U.S.A.* 111, 10642–10647. doi: 10.1073/pnas.1410761111
- Jablonka, E. (2013). Epigenetic inheritance and plasticity: the responsive germline. *Prog. Biophys. Mol. Biol.* 111, 99–107. doi: 10.1016/j.biombio.2012.08.014
- Jablonka, E., and Lamb, M. J. (1998). Epigenetic inheritance in evolution. *J. Evol. Biol.* 11, 159–183.
- Jacobsen, S. E., Sakai, H., Finnegan, E. J., Cao, X., and Meyerowitz, E. M. (2000). Ectopic hypermethylation of flower-specific genes in *Arabidopsis*. *Curr. Biol.* 10, 179–186. doi: 10.1016/S0960-9822(00)00324-9
- Janoušek, B., Šírký, J., and Vyskot, B. (1996). Epigenetic control of sexual phenotype in a dioecious plant, *Melandrium album*. *Mol. Gen. Genet.* 250, 483–490. doi: 10.1007/bf02174037
- Jeltsch, A. (2010). Molecular biology, phylogeny of methylomes. *Science* 328, 837–838. doi: 10.1126/science.1190738
- Johannes, F., Porcher, E., Teixeira, F. K., Saliba-Colombani, V., Simon, M., Agier, N., et al. (2009). Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genet.* 5:e1000530. doi: 10.1371/journal.pgen.1000530
- Kankel, M. W., Ramsey, D. E., Stokes, T. L., Flowers, S. K., Haag, J. R., Jeddeloh, J. A., et al. (2003). *Arabidopsis* MET1 cytosine methyltransferase mutants. *Genetics* 163, 1109–1122.
- Kawakatsu, T., Nery, J. R., Castanon, R., and Ecker, J. R. (2017). Dynamic DNA methylation reconfiguration during seed development and germination. *Genome Biol.* 18:171. doi: 10.1186/s13059-017-1251-x
- Kelliher, T., and Walbot, V. (2012). Hypoxia triggers meiotic fate acquisition in maize. *Science* 337, 345–348. doi: 10.1126/science.1220080
- Kinoshita, T., Miura, A., Choi, Y. H., Kinoshita, Y., Cao, X. F., Jacobsen, S. E., et al. (2004). One-way control of FWA imprinting in *Arabidopsis* endosperm by DNA methylation. *Science* 303, 521–523. doi: 10.1126/science.1089835
- Köhler, C., and Grossniklaus, U. (2002). Epigenetic inheritance of expression states in plant development: the role of Polycomb group proteins. *Curr. Opin. Cell Biol.* 14, 773–779. doi: 10.1016/S0955-0674(02)00394-0
- Kooke, R., Johannes, F., Wardenaar, R., Becker, F., Etcheverry, M., Colot, V., et al. (2015). Epigenetic basis of morphological variation and phenotypic plasticity in *Arabidopsis thaliana*. *Plant Cell* 27, 337–348. doi: 10.1105/tpc.114.133025
- Lai, Y. S., Zhang, X. H., Zhang, W., Shen, D., Wang, H. P., Xia, Y. D., et al. (2017). The association of changes in DNA methylation with temperature-dependent sex determination in cucumber. *J. Exp. Bot.* 68, 2899–2912. doi: 10.1093/jxb/erx144

- Lang, D., Ullrich, K. K., Murat, F., Fuchs, J., Jenkins, J., Haas, F. B., et al. (2018). The *Physcomitrella patens* chromosome scale assembly reveals moss genome structure evolution. *Plant J.* 93, 515–533. doi: 10.1111/tpj.13801
- Lang, Z. B., Wang, Y. H., Tang, K., Tang, D. G., Datsenko, T., Cheng, J. F., et al. (2017). Critical roles of DNA demethylation in the activation of ripening-induced genes and inhibition of ripening-repressed genes in tomato fruit. *Proc. Natl. Acad. Sci. U.S.A.* 114, E4511–E4519. doi: 10.1073/pnas.1705233114
- Law, J. A., and Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* 11, 204–220. doi: 10.1038/nrg2719
- Lees-Murdock, D. J., and Walsh, C. P. (2008). DNA methylation reprogramming in the germ line. *Epigenetics* 3, 5–13. doi: 10.4161/epi.3.1.5553
- Lei, M., La, H., Lu, K., Wang, P., Miki, D., Ren, Z., et al. (2014). *Arabidopsis* EDM2 promotes *IBM1* distal polyadenylation and regulates genome DNA methylation patterns. *Proc. Natl. Acad. Sci. U.S.A.* 111, 527–532. doi: 10.1073/pnas.1320106110
- Liu, J., Chatham, L., Aryal, R., Yu, Q., and Ming, R. (2018). Differential methylation and expression of HUA1 ortholog in three sex types of papaya. *Plant Sci.* 272, 99–106. doi: 10.1016/j.plantsci.2018.04.001
- Liu, S., Dunwell, T. L., Pfeifer, G. P., Dunwell, J. M., Ullah, I., and Wang, Y. (2013). Detection of oxidation products of 5-methyl-2'-deoxycytidine in *Arabidopsis* DNA. *PLoS One* 8:e84620. doi: 10.1371/journal.pone.0084620
- Malik, G., Dangwal, M., Kapoor, S., and Kapoor, M. (2012). Role of DNA methylation in growth and differentiation in *Physcomitrella patens* and characterization of cytosine DNA methyltransferases. *FEBS J.* 279, 4081–4094. doi: 10.1111/febs.12002
- Martin, A., Troadec, C., Boualem, A., Rajab, M., Fernandez, R., Morin, H., et al. (2009). A transposon-induced epigenetic change leads to sex determination in melon. *Nature* 461, 1135–1138. doi: 10.1038/nature08498
- Mathieu, O., Reinders, J., Caikovski, M., Smathajitt, C., and Paszkowski, J. (2007). Transgenerational stability of the *Arabidopsis* epigenome is coordinated by CG methylation. *Cell* 130, 851–862. doi: 10.1016/j.cell.2007.07.007
- Matzke, M. A., Kanno, T., and Matzke, A. J. (2015). RNA-directed DNA methylation: the evolution of a complex epigenetic pathway in flowering plants. *Annu. Rev. Plant Biol.* 66, 243–267. doi: 10.1146/annurev-arplant-043014-114633
- Mirouze, M., and Vitte, C. (2014). Transposable elements, a treasure trove to decipher epigenetic variation: insights from *Arabidopsis* and crop epigenomes. *J. Exp. Bot.* 65, 2801–2812. doi: 10.1093/jxb/eru120
- Mosquana, A., Katz, A., Decker, E. L., Rensing, S. A., Reski, R., and Ohad, N. (2009). Regulation of stem cell maintenance by the Polycomb protein FIE has been conserved during land plant evolution. *Development* 136, 2433–2444. doi: 10.1242/dev.035048
- Niederhuth, C. E., Bewick, A. J., Ji, L., Alabady, M. S., Kim, K. D., Li, Q., et al. (2016). Widespread natural variation of DNA methylation within angiosperms. *Genome Biol.* 17:194. doi: 10.1186/s13059-016-1059-0
- Niklas, K. J., Cobb, E. D., and Dunker, A. K. (2014). The number of cell types, information content, and the evolution of complex multicellularity. *Acta Soc. Bot. Pol.* 83, 337–347. doi: 10.5586/asbp.2014.034
- Nishiyama, T., Sakayama, H., de Vries, J., Buschmann, H., Saint-Marcoux, D., Ullrich, K. K., et al. (2018). The *Chara* genome: secondary complexity and implications for plant terrestrialization. *Cell* 174, 448–464. doi: 10.1016/j.cell.2018.06.033
- Okano, Y., Aono, N., Hiwatashi, Y., Murata, T., Nishiyama, T., Ishikawa, T., et al. (2009). A polycomb repressive complex 2 gene regulates apogamy and gives evolutionary insights into early land plant evolution. *Proc. Natl. Acad. Sci. U.S.A.* 106, 16321–16326. doi: 10.1073/pnas.0906997106
- Pennell, R. I., Cronk, Q. C. B., Forsberg, S., Stohr, C., Snogerup, L., Kjellbom, P., et al. (1995). Cell-context signaling. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 350, 87–93. doi: 10.1098/rstb.1995.0142
- Pigliucci, M., Murren, C. J., and Schlichting, C. D. (2006). Phenotypic plasticity and evolution by genetic assimilation. *J. Exp. Biol.* 209, 2362–2367. doi: 10.1242/jeb.02070
- Raj, S., Bräutigam, K., Hamanishi, E. T., Wilkins, O., Thomas, B. R., Schroeder, W., et al. (2011). Clone history shapes *Populus* drought responses. *Proc. Natl. Acad. Sci. U.S.A.* 108, 12521–12526. doi: 10.1073/pnas.1103341108
- Rodrigues, J. A., and Zilberman, D. (2015). Evolution and function of genomic imprinting in plants. *Genes Dev.* 29, 2517–2531. doi: 10.1101/gad.269902.115
- Roudier, F., Ahmed, I., Berard, C., Sarazin, A., Mary-Huard, T., Cortijo, S., et al. (2011). Integrative epigenomic mapping defines four main chromatin states in *Arabidopsis*. *EMBO J.* 30, 1928–1938. doi: 10.1038/emboj.2011.103
- Roudier, F., Teixeira, F. K., and Colot, V. (2009). Chromatin indexing in *Arabidopsis*: an epigenomic tale of tails and more. *Trends Genet.* 25, 511–517. doi: 10.1016/j.tig.2009.09.013
- Roux, F., Colome-Tatche, M., Edelist, C., Wardenaar, R., Guerche, P., Hospital, F., et al. (2011). Genome-wide epigenetic perturbation jump-starts patterns of heritable variation found in nature. *Genetics* 188, 1015–1017. doi: 10.1534/genetics.111.128744
- Rudall, P. J., and Bateman, R. M. (2003). Evolutionary change in flowers and inflorescences: evidence from naturally occurring terata. *Trends Plant Sci.* 8, 76–82. doi: 10.1016/s1360-1385(02)00026-2
- Ruhfel, B. R., Gitzendanner, M. A., Soltis, P. S., Soltis, D. E., and Burleigh, J. G. (2014). From algae to angiosperms—inferring the phylogeny of green plants (*Viridiplantae*) from 360 plastid genomes. *BMC Evol. Biol.* 14:23. doi: 10.1186/1471-2148-14-23
- Saze, H., and Kakutani, T. (2007). Heritable epigenetic mutation of a transposon-flanked *Arabidopsis* gene due to lack of the chromatin-remodeling factor DDM1. *EMBO J.* 26, 3641–3652. doi: 10.1038/sj.emboj.7601788
- Schmid, M. W., Giraldo-Fonseca, A., Rövekamp, M., Smetanin, D., Bowman, J. L., and Grossniklaus, U. (2018). Extensive epigenetic reprogramming during the life cycle of *Marchantia polymorpha*. *Genome Biol.* 19:9. doi: 10.1186/s13059-017-1383-z
- Scoville, A. G., Barnett, L. L., Bodbyl-Roels, S., Kelly, J. K., and Hileman, L. C. (2011). Differential regulation of a MYB transcription factor is correlated with transgenerational epigenetic inheritance of trichome density in *Mimulus guttatus*. *New Phytol.* 191, 251–263. doi: 10.1111/j.1469-8137.2011.03656.x
- Sigman, M. J., and Slotkin, R. K. (2016). The first rule of plant transposable element silencing: location, location, location. *Plant Cell* 28, 304–313. doi: 10.1105/tpc.15.00869
- Slotkin, R. K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* 8, 272–285. doi: 10.1038/nrg2072
- Slotkin, R. K., Vaughn, M., Borges, F., Tanurdzic, M., Becker, J. D., Feijo, J. A., et al. (2009). Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* 136, 461–472. doi: 10.1016/j.cell.2008.12.038
- Song, Y., Ma, K., Ci, D., Chen, Q., Tian, J., and Zhang, D. (2013a). Sexual dimorphic floral development in dioecious plants revealed by transcriptome, phytohormone, and DNA methylation analysis in *Populus tomentosa*. *Plant Mol. Biol.* 83, 559–576. doi: 10.1007/s11103-013-0108-2
- Song, Y., Ma, K., Ci, D., Zhang, Z., and Zhang, D. (2013b). Sexual dimorphism floral MicroRNA profiling and target gene expression in andromonoecious poplar (*Populus tomentosa*). *PLoS One* 8:e62681. doi: 10.1371/journal.pone.0062681
- Soppe, W. J., Jacobsen, S. E., Alonso-Blanco, C., Jackson, J. P., Kakutani, T., Koornneef, M., et al. (2000). The late flowering phenotype of *fwa* mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Mol. Cell* 6, 791–802. doi: 10.1016/S1097-2765(05)00090-0
- Stroud, H., Greenberg, M. V., Feng, S., Bernatavichute, Y. V., and Jacobsen, S. E. (2013). Comprehensive analysis of silencing mutants reveals complex regulation of the *Arabidopsis* methylome. *Cell* 152, 352–364. doi: 10.1016/j.cell.2012.10.054
- Sulston, J. E., and Horvitz, H. R. (1977). Post-embryonic cell lineages of nematode, *Caenorhabditis elegans*. *Dev. Biol.* 56, 110–156. doi: 10.1016/0012-1606(77)90158-0
- Takuno, S., and Gaut, B. S. (2012). Body-methylated genes in *Arabidopsis thaliana* are functionally important and evolve slowly. *Mol. Biol. Evol.* 29, 219–227. doi: 10.1093/molbev/msr188
- Takuno, S., and Gaut, B. S. (2013). Gene body methylation is conserved between plant orthologs and is of evolutionary consequence. *Proc. Natl. Acad. Sci. U.S.A.* 110, 1797–1802. doi: 10.1073/pnas.1215380110
- Takuno, S., Ran, J. H., and Gaut, B. S. (2016). Evolutionary patterns of genic DNA methylation vary across land plants. *Nat. Plants* 2:15222. doi: 10.1038/nplants.2015.222
- Tennessen, J. A., Govindarajulu, R., Liston, A., and Ashman, T. L. (2016). Homomorphic ZW chromosomes in a wild strawberry show distinctive

- recombination heterogeneity but a small sex-determining region. *New Phytol.* 211, 1412–1423. doi: 10.1111/nph.13983
- Timme, R. E., Bachvaroff, T. R., and Delwiche, C. F. (2012). Broad phylogenomic sampling and the sister lineage of land plants. *PLoS One* 7:e29696. doi: 10.1371/journal.pone.0029696
- To, T. K., Nakaminami, K., Kim, J. M., Morosawa, T., Ishida, J., Tanaka, M., et al. (2011). *Arabidopsis HDA6* is required for freezing tolerance. *Biochem. Biophys. Res. Commun.* 406, 414–419. doi: 10.1016/j.bbrc.2011.02.058
- Tricker, P. J., Gibbings, J. G., Rodríguez, López CM, Hadley, P., and Wilkinson, M. J. (2012). Low relative humidity triggers RNA-directed de novo DNA methylation and suppression of genes controlling stomatal development. *J. Exp. Bot.* 63, 3799–3813. doi: 10.1093/jxb/ers076
- Tricker, P. J., López, C. M., Gibbings, G., Hadley, P., and Wilkinson, M. J. (2013b). Transgenerational, dynamic methylation of stomata genes in response to low relative humidity. *Int. J. Mol. Sci.* 14, 6674–6689. doi: 10.3390/ijms14046674
- Tricker, P. J., Rodríguez López, C., Hadley, P., Wagstaff, C., and Wilkinson, M. (2013a). Pre-conditioning the epigenetic response to high vapor pressure deficit increases the drought tolerance of *Arabidopsis thaliana*. *Plant Signal. Behav.* 8:e25974. doi: 10.4161/psb.25974
- Turmel, M., Otis, C., and Lemieux, C. (2006). The chloroplast genome sequence of *Chara vulgaris* sheds new light into the closest green algal relatives of land plants. *Mol. Biol. Evol.* 23, 1324–1338. doi: 10.1093/molbev/msk018
- van der Graaf, A., Wardenaar, R., Neumann, D. A., Taudt, A., Shaw, R. G., Jansen, R. C., et al. (2015). Rate, spectrum, and evolutionary dynamics of spontaneous epimutations. *Proc. Natl. Acad. Sci. U.S.A.* 112, 6676–6681. doi: 10.1073/pnas.1424254112
- Vidalis, A., Živković, D., Wardenaar, R., Roquis, D., Tellier, A., and Johannes, F. (2016). Methylome evolution in plants. *Genome Biol.* 17:264. doi: 10.1186/s13059-016-1127-5
- Whipple, C. (2012). Defining the plant germ line-nature or nurture? *Science* 337, 301–302. doi: 10.1126/science.1224362
- Wickett, N. J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., et al. (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
- Wood, C. C., Robertson, M., Tanner, G., Peacock, W. J., Dennis, E. S., and Helliwell, C. A. (2006). The *Arabidopsis thaliana* vernalization response requires a polycomb-like protein complex that also includes VERNALIZATION INSENSITIVE 3. *Proc. Natl. Acad. Sci. U.S.A.* 103, 14631–14636. doi: 10.1073/pnas.0606385103
- Xiao, J., Jin, R., and Wagner, D. (2017). Developmental transitions: integrating environmental cues with hormonal signaling in the chromatin landscape in plants. *Genome Biol.* 18:88. doi: 10.1186/s13059-017-1228-9
- Xiao, W., Custard, K. D., Brown, R. C., Lemmon, B. E., Harada, J. J., Goldberg, R. B., et al. (2006). DNA methylation is critical for *Arabidopsis* embryogenesis and seed viability. *Plant Cell* 18, 805–814. doi: 10.1105/tpc.105.038836
- Yaari, R., Noy-Malka, C., Wiedemann, G., Gershovitz, N. A., Reski, R., Katz, A., et al. (2015). DNA METHYLTRANSFERASE 1 is involved in (m)CG and (m)CCG DNA methylation and is essential for sporophyte development in *Physcomitrella patens*. *Plant Mol. Biol.* 88, 387–400. doi: 10.1007/s11103-015-0328-8
- Yamamuro, C., Miki, D., Zheng, Z. M., Ma, J., Wang, J., Yang, Z. B., et al. (2014). Overproduction of stomatal lineage cells in *Arabidopsis* mutants defective in active DNA demethylation. *Nat. Commun.* 5:4062. doi: 10.1038/ncomms5062
- Zemach, A., McDaniel, I. E., Silva, P., and Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 328, 916–919. doi: 10.1126/science.1186366
- Zemach, A., and Zilberman, D. (2010). Evolution of eukaryotic DNA methylation and the pursuit of safer sex. *Curr. Biol.* 20, R780–R785. doi: 10.1016/j.cub.2010.07.007
- Zhang, G., Huang, H., Liu, D., Cheng, Y., Liu, X., Zhang, W., et al. (2015). N6-methyladenine DNA modification in *Drosophila*. *Cell* 161, 893–906. doi: 10.1016/j.cell.2015.04.018
- Zhang, W., Wang, X., Yu, Q., Ming, R., and Jiang, J. (2008). DNA methylation and heterochromatinization in the male-specific region of the primitive Y chromosome of papaya. *Genome Res.* 18, 1938–1943. doi: 10.1101/gr.078808.108
- Zhang, Y., Harris, C. J., Liu, Q., Liu, W., Ausin, I., Long, Y., et al. (2018). Large-scale comparative epigenomics reveals hierarchical regulation of non-CG methylation in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 115, E1069–E1074. doi: 10.1073/pnas.1716300115
- Zheng, X. G., Chen, L., Xia, H., Wei, H. B., Lou, Q. J., Li, M. S., et al. (2017). Transgenerational epimutations induced by multi-generation drought imposition mediate rice plant's adaptation to drought condition. *Sci. Rep.* 7:39843. doi: 10.1038/srep39843
- Zhou, C., Wang, C., Liu, H., Zhou, Q., Liu, Q., Guo, Y., et al. (2018). Identification and analysis of adenine N6-methylation sites in the rice genome. *Nat. Plants* 4, 554–563. doi: 10.1038/s41477-018-0214-x
- Zilberman, D. (2017). An evolutionary case for functional gene body methylation in plants and animals. *Genome Biol.* 18:87. doi: 10.1186/s13059-017-1230-2

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 © 2018 Bräutigam and Cronk. 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 Times They Are A-Changin': Heterochrony in Plant Development and Evolution

Manuel Buendía-Monreal* and C. Stewart Gillmor

Laboratorio Nacional de Genómica para la Biodiversidad (Langebio), Unidad de Genómica Avanzada, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (CINVESTAV-IPN), Guanajuato, Mexico

OPEN ACCESS

Edited by:

Natalia Pabón-Mora,
Universidad de Antioquia, Colombia

Reviewed by:

Jianfei Zhao,
University of Pennsylvania,
United States

Jill Christine Preston,
The University of Vermont,
United States

*Correspondence:

Manuel Buendía-Monreal
manuel.buendia@cinvestav.mx

Specialty section:

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

Received: 30 June 2018

Accepted: 27 August 2018

Published: 18 September 2018

Citation:

Buendía-Monreal M and Gillmor CS
(2018) The Times They Are
A-Changin': Heterochrony in Plant
Development and Evolution.
Front. Plant Sci. 9:1349.
doi: 10.3389/fpls.2018.01349

Alterations in the timing of developmental programs during evolution, that lead to changes in the shape, or size of organs, are known as heterochrony. Heterochrony has been widely studied in animals, but has often been neglected in plants. During plant evolution, heterochronic shifts have played a key role in the origin and diversification of leaves, roots, flowers, and fruits. Heterochrony that results in a juvenile or simpler outcome is known as paedomorphosis, while an adult or more complex outcome is called peramorphosis. Mechanisms that alter developmental timing at the cellular level affect cell proliferation or differentiation, while those acting at the tissue or organismal level change endogenous aging pathways, morphogen signaling, and metabolism. We believe that wider consideration of heterochrony in the context of evolution will contribute to a better understanding of plant development.

Keywords: heterochrony, developmental timing, plant development, plant evolution, cell cycle, miR156

THE DIFFERENT TYPES OF HETEROCHRONY

In the 1870s, Ernst Haeckel identified temporal and spatial changes in development in a descendant relative to its ancestor as the two mechanisms most important for evolution (Haeckel, 1875). Haeckel named spatial changes heterotopy, and temporal changes heterochrony.

However, the meaning of heterochrony has changed since Haeckel first coined the term. Haeckel used the term heterochrony to refer to deviations from his well-known "Biogenetic Law," which states that the sequence of developmental events (ontogeny) largely recapitulates the sequence of events in the evolutionary history of the species (phylogeny) (Haeckel, 1875). Thus, heterochrony originally referred to a change in the timing of appearance of a feature in a developmental sequence of an organism, relative to the sequence that occurred in the organism's phylogeny. In the middle of the 20th century, De Beer (1951) uncoupled heterochrony from recapitulation. He used heterochrony to denote differences in the timing of developmental events when comparing two related species, to explain how heterochrony could generate diversity among organisms. Gould (1977) re-associated the concept of heterochrony to recapitulation, defining heterochrony as "changes in the relative time of appearance and rate of development for characters already present in ancestors," emphasizing changes in relative size and shape, rather than in the timing of developmental events, to detect heterochrony. At the turn of the 21st century, Smith proposed that it would be more useful to define two types of heterochrony: 'growth heterochrony,' which emphasizes changes in final size and shape; and 'sequence heterochrony,' which is closer to the original usage of Haeckel and de Beer, and allows explanation of phenotypic variation by changes in the timing of developmental events (Smith, 2002, 2003; Keyte and Smith, 2014).

Sequence heterochrony (hereafter referred to as ‘heterochrony’) can be classified in two categories: paedomorphosis and peramorphosis. When compared to ancestral development, paedomorphosis results in a juvenile or simple outcome, whereas peramorphosis results in an adult or more complex phenotype. Each of these two categories of heterochrony can result from variation in timing of the onset, offset or rate of a developmental process, as proposed by Alberch et al. (1979). This variation can result in 6 different types of heterochrony (**Figure 1**). Paedomorphosis can result from the precocious end of a developmental process (progenesis), from a delayed start of the process (post-displacement), or from a slower rate of development (neoteny). Peramorphosis is the result of an extended period of development due to a later termination (hypermorphosis) or an earlier onset (pre-displacement), or of a higher rate of development (acceleration) (**Figure 1A**). Hypothetical examples for peramorphosis and paedomorphosis in plant embryogenesis and vegetative development are shown in **Figures 1B,C**.

While these classifications are useful to illustrate changes in developmental timing, when considering real life examples, it is often difficult to distinguish between different kinds of heterochrony. In many cases, evolution can result in distinct types of heterochrony, each occurring at a discrete stage of a developmental process (reviewed in plants by Li and Johnston, 2000).

HETEROCHRONY IN THE EVOLUTION AND DIVERSIFICATION OF PLANTS

Land plants (embryophytes) have undergone many morphological innovations since their emergence in the mid-Ordovician period, about 470 million years ago. Early diverging lineages originated over a period of more than 100 million years, during the Silurian and early Devonian periods (Kenrick and Crane, 1997; Pires and Dolan, 2012; Harrison and Morris, 2018). Embryophytes evolved from a freshwater algae ancestor that was related to the extant charophyte groups Charales, Coleochaetales, and Zygnematales. The transition to growth on land, as observed in bryophytes (hornworts, mosses, and liverworts), involved the origin of spores, alternation of gametophyte and sporophyte generations, uniaxial forms, and three-dimensional growth. Innovations of terrestrial vascular plants (lycophytes, monilophytes, and spermatophytes) included bifurcation, indeterminacy, sporophytic dominance, axillary branching, and the formation of meristems, leaves, and roots (Harrison, 2017).

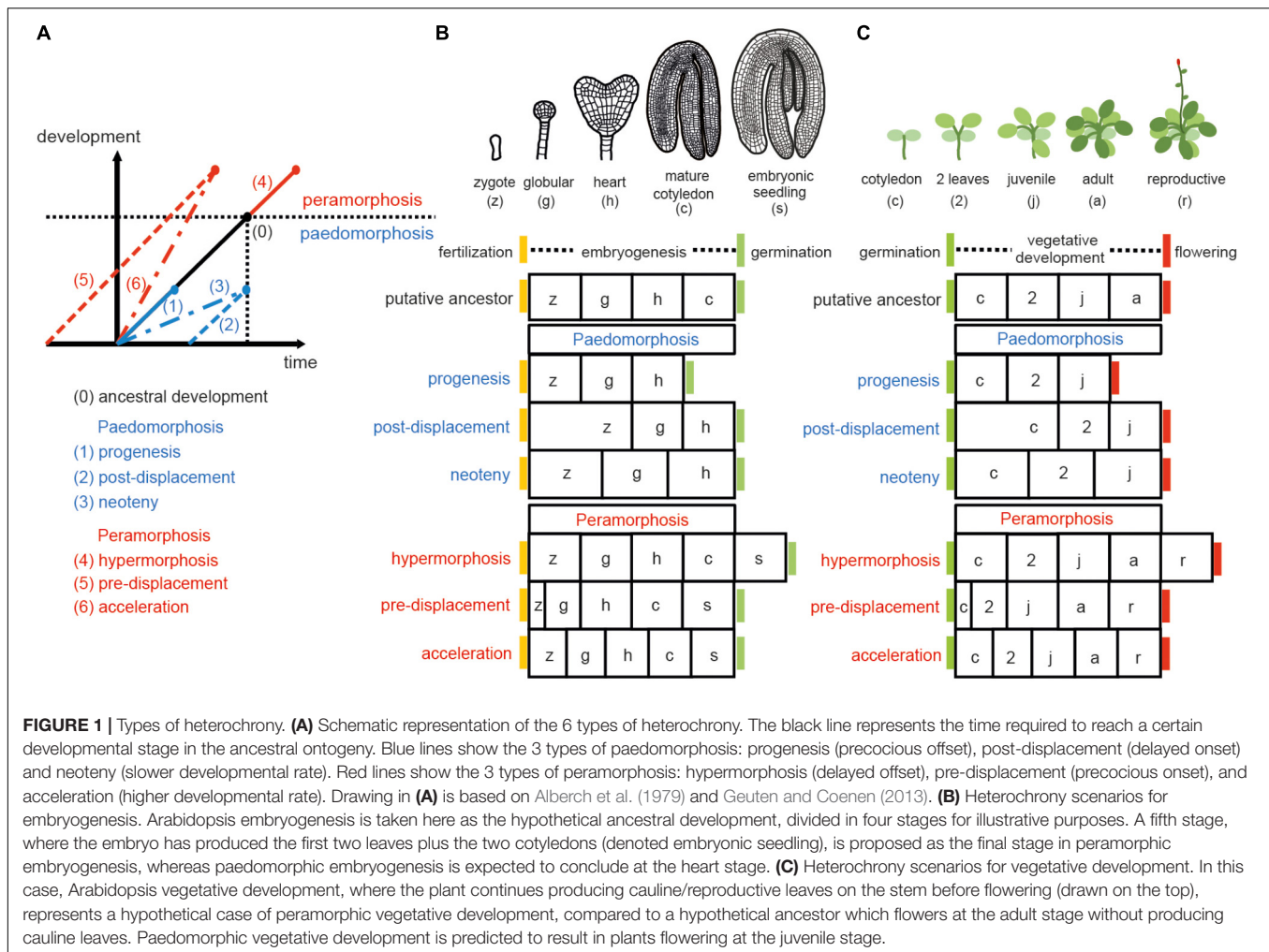
Several of the evolutionary steps above have been linked to heterochrony. For the origin of sporogenesis, bryophyte data suggest that spores were produced directly from zygotes in a process involving precocious cytokinesis, acceleration of meiosis and delayed wall deposition from the zygote to the meiospores (Brown and Lemmon, 2011). The branched sporophyte (polisporangiophyte) has been hypothesized to have evolved by extended vegetative growth of the apical cell. This

longer period of vegetative growth was proposed to result in a prolonged embryonic axis, shoot branching, and a delay in the transition to reproductive growth, producing the sporangium (Rothwell et al., 2014; Tomescu et al., 2014). Additional studies of plant fossils should provide more evidence for ancient plant morphologies, which would allow comparison between contemporary and extinct forms (Rothwell et al., 2014).

In extant plants, heterochronic changes have been identified in gametophyte development, embryogenesis, vegetative development, shoot maturation, and floral morphogenesis. The female gametophyte of *Gnetum* is structurally divergent from other plants because of differences in the timing of fertilization and somatic development. Temporal alterations in cell cycle progression have contributed to diversified temporal patterns of spermatogenesis and gamete fusion during fertilization (Friedman, 1999; Tian et al., 2005). In a case of progenesis, fertilization occurs at a free nuclear stage of somatic development, a juvenile stage compared to the ancestral somatic ontogeny, precluding the differentiation of egg cells (Friedman and Carmichael, 1998). The apomictic development of *Boechera* ovules has been associated with heterochronic gene expression patterns compared to non-apomictic (sexual) ovules (Sharbel et al., 2010). The development of Rafflesiaceae, a holoparasitic plant family which infects grapevines, shows two heterochronic shifts: an arrest at the proembryonic stage, which can be considered an example of neoteny, and acceleration of the transition from the undifferentiated endophyte to flowering, skipping vegetative shoot maturation (Nikolov et al., 2014).

A Quantitative Trait Loci (QTL) analysis comparing *Eucalyptus globulus* populations with precocious vegetative phase change and populations in which vegetative phase change is delayed several years identified the expression of the microRNA EglMIR156.5 as responsible for heterochronic variation in vegetative phase change in *E. globulus* (Hudson et al., 2014). Another QTL analysis concluded that heterochrony underlies natural variation in *Cardamine hirsuta* leaf form (Cartolano et al., 2015). QTL mapping determined that the effect is caused by *cis*-regulatory variation in the floral repressor *ChFLC* such that populations with low-expressing *ChFLC* alleles show both early flowering and accelerated acquisition of adult leaf traits, particularly increased leaflet number. Morphometric and QTL analyses have determined that heterochronic mutations contribute to natural variation in *Antirrhinum* and to grapevine heteroblasty (Costa et al., 2012; Chitwood et al., 2016). A Principal Component Analysis (PCA) of the ontogenetic trajectories of leaf form among the three genera of marsileaceous ferns (*Marsilea*, *Regnellidium*, and *Pilularia*) suggested that they show a paedomorphic phenotype, compared to the more complex ancestral development, caused by accelerated growth rate and early termination at a simplified leaf form (Pryer and Hearn, 2009).

Evolutionary diversity in inflorescence architecture of the Solanaceae is modulated by heterochronic shifts in the acquisition of floral fate (Lippman et al., 2008; Park et al., 2012). A comparison of transcriptomes of meristem maturation from five domesticated and wild Solanaceae species revealed a peak of expression divergence, resembling the “inverse hourglass” model



for animal embryogenesis, which states that a mid-development period of divergence drives morphological variation (Lemmon et al., 2016). In grasses, a delay in the shoot meristem (SM) to floral meristem (FM) transition results in more complex panicles (Kyojuka et al., 2014). Poplars (*Populus* sp.) and willows (*Salix* sp.) bear compact unisexual inflorescences known as “catkins,” which have been proposed to be evolved from a simplification of the panicle form by an early SM to FM transition (Cronk et al., 2015). Heterochronic changes have also contributed to natural variation in flowering time and shoot architecture among *Mimulus guttatus* populations (Baker and Diggle, 2011).

Morphological diversity of the perianth in Dipsacoidae is caused by heterochronic changes in organ initiation, specifically in the number of sepals (Naghiloo and Claßen-Bockhoff, 2017). The great shape diversity of sepals among Iris species is due more to heterochrony than to heterotopic changes (Guo, 2015). A study in Brassicaceae showed that evolution of corolla monosymmetry from the polysymmetrical ancestral flower involved a heterochronic shift in the expression of *CYC2* genes (a clade of *TCP* transcription factors) from early adaxial expression in ancestral floral meristems, to a later adaxial expression in petal development (Busch et al.,

2012). Heterochronic, but not heterotopic, *CYC2* expression has also been associated with a loss of papillate conical cells in petals and a shift to bird-pollination system in *Lotus* (Ojeda et al., 2017). A paedomorphic morphology, in which the flowers hold mature pollen in unopened bud-like structures, led to specialized pollination in a clade of Madagascar vines (Euphorbiaceae) (Armbruster et al., 2013). The evolution of cleistogamous capitulum from a chasmogamous ancestral state is a classic example of paedomorphosis, since the cleistogamous shape shows juvenile traits (Lord and Hill, 1987). Cleistogamy in Asteraceae specifically evolved by pre-displacement and progenesis of floral development, as well as neoteny of all whorls other than the gynoecium (Porras and Muñoz, 2000). The diversity of floral morphologies within *Jaltomata*, a Solanaceae genus, is due to hypermorphosis and acceleration of some corolla traits (Kostyun et al., 2017). The diversity of Azorean butterfly orchids is also caused by floral heterochronic shifts (Bateman et al., 2014). Recently, Ronse de Craene (2018) emphasized the importance of heterochrony in three developmental processes: phyllotaxis, the development of common stamen-petal primordia and obdiplostemony, linking changes in the growth rate with delayed organ initiation. Heterochronic growth rates of

the perianth and style, and early hypanthium elongation, are responsible for the great species diversity within the morphologically homogeneous *Eugenia* genus (Vasconcelos et al., 2018). Finally, heterochronic expression of the *fw2.2* allele, which affects cell division in early fruit development, is responsible for natural variation in tomato fruit size (Cong et al., 2002).

STUDYING HETEROCHRONIC MUTANTS TO ELUCIDATE GENETIC CONTROL OF TIMING

The study of mutants affected in developmental timing has shed light on genetic pathways controlling morphogenesis and developmental transitions. Heterochrony can be caused by earlier or later activation or repression of these pathways.

leafy cotyledon (*lec*), *dicer-like1* (*dcl1*), and *extra cotyledon* (*xtc*) mutants of *Arabidopsis thaliana* represent heterochronic phenotypes that have helped to define seed maturation programs. *lec* mutants produce cotyledons with features of leaf identity (Meinke, 1992), a clear example of homeosis: the replacement of one structure by another. However, it is often difficult to distinguish between homeosis and heterochrony, since homeosis can be the result of both heterochrony and heterotopy (Li and Johnston, 2000; Geuten and Coenen, 2013). During late embryogenesis, *LEC2* promotes seed maturation and represses postembryonic identity (Stone et al., 2008). Like *lec* mutants, *dcl1* mutants show peramorphic phenotypes during embryogenesis, as chloroplast development and seed storage protein gene expression occur earlier than in wild type embryos. *DCL1* is required for biogenesis of microRNAs, which repress seed maturation through the master regulators *LEC2* and *FUSCA3* (Nordine and Bartel, 2010; Willmann et al., 2011). The *xtc1*, *xtc2*, and *altered meristem programming1* (*amp1*) mutants show a homeotic phenotype where the first one or two leaves are transformed into cotyledons (Conway and Poethig, 1997). In these three mutants, the globular to heart transition is delayed, causing an enlarged shoot meristem, which leads to extra organ formation during embryogenesis. This phenotype can be interpreted as hypermorphosis, a type of peramorphosis, since embryo development is extended to a more developed shape (like the peramorphic scenario in **Figure 1B**). However, if vegetative development is chosen as a point of reference, *lec* mutants would represent a case of pre-displacement (peramorphosis), and *xtc1*, *xtc2*, and *amp1* would represent post-displacement (paedomorphosis) in the acquisition of leaf identity. AMP1 is required for miRNA-mediated translational repression on the ER membrane (Li et al., 2013).

Genetic regulation of the juvenile to adult transition, also called vegetative phase change, has been studied using both paedomorphic and peramorphic mutants. Screenings for mutants showing an early adult (peramorphic) phenotype produced alleles of genes related to small RNA biogenesis like *zippy/ago7*, *sgs3*, and *rdr6* (Hunter et al., 2003; Peragine et al., 2004), while mutants with a late adult (paedomorphic) phenotype are due to increased miRNA levels (Gillmor et al., 2014; Xu M. et al., 2016, 2018; Guo et al., 2017). The genetic basis for

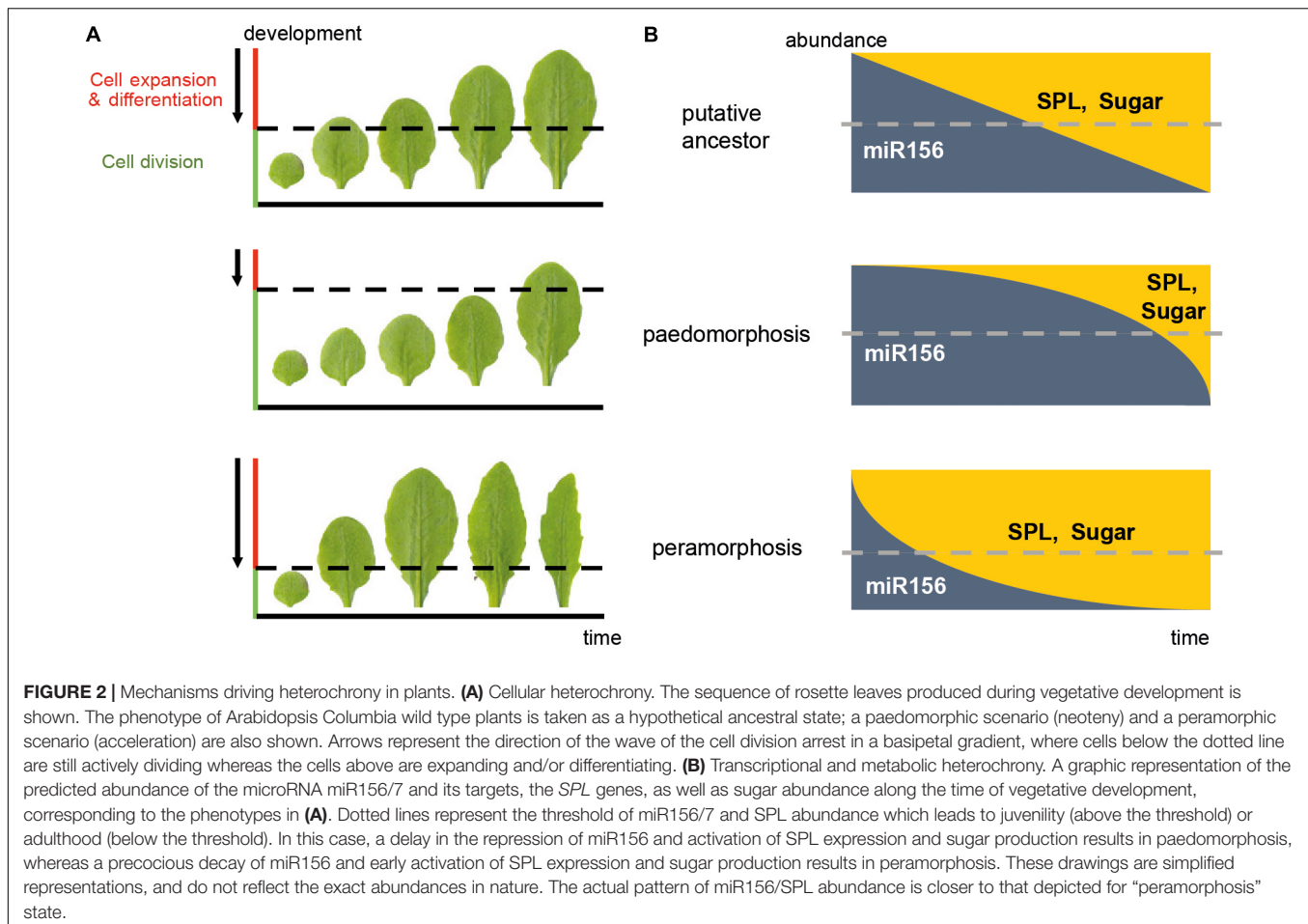
these phenotypes is explained by altered expression of the closely related microRNAs miR156 and miR157, the main regulators of vegetative phase change that act by repressing *SPL* transcription factors at both transcriptional and translational levels, in a threshold-dependent manner (Wu and Poethig, 2006; Chuck et al., 2007; Poethig, 2013; He et al., 2018). Both *MIR156* and *SPL* gene families are directly regulated by epigenetic marks: *MIR156* is activated by H3K4me3 and H3K4ac, which are promoted by the SWR1-C complex and the chromatin remodeler BRAHMA, and is repressed by H3K27me3, which is promoted by the Polycomb proteins SWINGER and CURLY LEAF and by the chromatin remodeler PICKLE (Xu M. et al., 2016; Xu Y. et al., 2016; Xu M. et al., 2018). *SPL* genes are activated by histone acetylation mediated by the SAGA-like complex and repressed by H2AUb mediated by the Polycomb proteins RING1A and RING1B (Kim et al., 2015; Li et al., 2017). The study of plants with loss- or gain-of function of *SPL* genes has also defined an endogenous flowering pathway in which some *SPL* genes promote the expression of miR172, which in turn promotes flowering by repressing the *APETALA2* family flowering repressors (Wang et al., 2009; Wu et al., 2009; Huijser and Schmid, 2011).

Besides major developmental transitions, organ morphogenesis can also be affected by heterochronic activation of genetic regulators. For instance, early germination of *Brassica rapa* embryos results in organs with mosaics of cotyledon and leaf identity (Fernandez, 1997), and differential temporal expression of the class II of *TEOSINTE BRANCHED1/CYCLOIDEA/PROLIFERATION CELL FACTOR* (*TCP*) genes results in leaves with different size and shape (Efroni et al., 2008).

MECHANISMS DRIVING CHANGES IN DEVELOPMENTAL TIMING IN PLANTS (TRANSCRIPTIONAL, METABOLIC, AND CELLULAR HETEROCHRONY)

Transcriptional heterochrony refers to a change in the timing of activation or repression of gene expression, and is often caused by changes in *cis*-regulatory gene regions (Pham et al., 2017). Transcriptional heterochrony in the genetic pathways mentioned above is a common way of producing heterochronic phenotypes. Metabolic control of pathways regulating developmental transitions (referred to here as “metabolic heterochrony”), and temporal control of cell proliferation, cell expansion and cell differentiation (referred to here as “cellular heterochrony”) are other mechanisms driving heterochrony in plants (**Figure 2**). Transcriptional, metabolic and cellular processes are interconnected, so the molecular origin of heterochrony can be due to a combination of mechanisms.

Metabolic heterochrony can be influenced by hormones, sugars, and redox signals (Jia et al., 2017). Differential biosynthesis, transport and perception of hormones such as auxin, jasmonic acid (JA), gibberellin and abscisic acid influence heterochrony by controlling regulators of developmental processes. For instance, mutants in the *AUXIN RESPONSE*



FACTORS (ARF) *ARF3* and *ARF4* delay the adult transition (Fahlgren et al., 2006; Hunter et al., 2006), and auxin homeostasis controls the transition from floral stem cell maintenance to gynoecium formation (Yamaguchi et al., 2017). Exogenous JA can delay the adult transition by postponing the decline of miR156 expression (Beydler et al., 2016), and gibberellin accelerates flowering by releasing *SPL* genes from repression by DELLA proteins (Yu et al., 2012). Nutritional status has been associated with the control of vegetative phase change since the early 20th century (Goebel, 1908). Sugar produced by photosynthesis is necessary for the acquisition of adult traits and is partially responsible for the decrease in miR156 expression in late vegetative development (Yang et al., 2013; Yu et al., 2013; Buendía-Monreal and Gillmor, 2017). *HEXOKINASE1* (*HXK1*) and Trehalose-6-phosphate (T6P) are important for the sugar-mediated repression of miR156, thereby promoting vegetative phase change and flowering (Wahl et al., 2013; Yang et al., 2013).

At the cellular level, organogenesis consists of a sequence of three stages: the establishment of polarity, cell proliferation, and cell expansion (Walcher-Chevillet and Kramer, 2016). The timing of initiation and termination of these stages is crucial for the size and shape of organs, and heterochrony in this sequence results in diversification of organ size and shape

(Figure 2). The transition from cell proliferation to cell expansion and differentiation requires coordination between the cell cycle and cell growth (Sablowski and Carnier Dornelas, 2014). In leaves of model plants, this transition moves as a basipetal wave of cell cycle arrest that begins at the distal part of the primordium and moves to the base. Cells behind the mitotic arrest front become highly vacuolated and begin to expand (Donnelly et al., 1999; Czesnick and Lenhard, 2015). However, other plant species can show diffuse growth, and acropetal or bidirectional cell cycle arrest gradients (Das Gupta and Nath, 2015). The acquisition of photosynthetic capacity is required for the shift from cell division to cell expansion (Andriankaja et al., 2012). This shift correlates with the role of sugar in promoting the Target of Rapamycin (TOR) pathway and repressing the Sucrose-non-fermenting1-related kinase 1 (SnRK1): TOR and T6P induce cell expansion by promoting macromolecular synthesis, whereas SnRK1 promotes catabolism (Tsai and Gazzarrini, 2014; Sablowski, 2016). Two microRNAs play opposite roles in this cellular shift: miR319 represses the expression of class II TCP factors, which are inhibitors of cell proliferation, whereas miR396 restricts the expression of GROWTH REGULATING FACTORS (GRFs), which delay differentiation (Das Gupta and Nath, 2015; Maugarny-Calés and Laufs, 2018). The transition from an indeterminate shoot apical meristem to a determinate

floral meristem also involves temporal regulation of the cellular identity. The timing of *AGAMOUS* activation of *KNUCKLES*, which in turn represses *WUSCHEL*, defines the temporal window of indeterminacy and consequently the size and the number of organs (Sun et al., 2014).

CONCLUSION

An understanding of temporal regulation of plant development is necessary to better appreciate the diversity of plant forms that we see in nature, to explain plant morphological evolution, and to manipulate plant architecture for the benefit of agriculture.

REFERENCES

- Alberch, P., Gould, S., Oster, G., and Wake, D. (1979). Size and shape in ontogeny and phylogeny. *Paleobiology* 5, 296–317. doi: 10.1017/S0094837300006588
- Andriankaja, M., Dhondt, S., DeBodt, S., Vanhaeren, H., Coppens, F., DeMilde, L., et al. (2012). Exit from proliferation during leaf development in *Arabidopsis thaliana*: a not-so-gradual process. *Dev. Cell* 22, 64–78. doi: 10.1016/j.devcel.2011.11.011
- Armbruster, W. S., Lee, J., Edwards, M. E., and Baldwin, B. G. (2013). Floral paedomorphy leads to secondary specialization in pollination of Madagascar *Dalechampia* (Euphorbiaceae). *Evolution* 67, 1196–1203. doi: 10.1111/evo.12002
- Baker, R. L., and Diggle, P. K. (2011). Node-specific branching and heterochronic changes underlie population-level differences in *Mimulus guttatus* (Phrymaceae) shoot architecture. *Am. J. Bot.* 98, 1924–1934. doi: 10.3732/ajb.1100098
- Bateman, R. M., Rudall, P. J., Bidartondo, M. I., Cozzolino, S., Tranchida-Lombardo, V., Carine, M. A., et al. (2014). Speciation via floral heterochrony and presumed mycorrhizal host switching of endemic butterfly orchids on the Azorean archipelago. *Am. J. Bot.* 101, 979–1001. doi: 10.3732/ajb.1300430
- Beydler, B. D., Osadchuk, K., Cheng, C.-L., Manak, J. R., and Irish, E. E. (2016). The juvenile phase of maize sees upregulation of stress-response genes and is extended by exogenous JA. *Plant Physiol.* 171, 2648–2658. doi: 10.1104/pp.15.01707
- Brown, R. C., and Lemmon, B. E. (2011). Spores before sporophytes: hypothesizing the origin of spore- genesis at the algal – plant transition. *New Phytol.* 190, 875–881. doi: 10.1111/j.1469-8137.2011.03709.x
- Buendía-Monreal, M., and Gillmor, C. S. (2017). Convergent repression of miR156 by sugar and the CDK8 module of *Arabidopsis* Mediator. *Dev. Biol.* 423, 19–23. doi: 10.1016/j.ydbio.2017.01.007
- Busch, A., Horn, S., Mühlhausen, A., Mummenhoff, K., and Zachgo, S. (2012). Corolla monosymmetry: evolution of a morphological novelty in the Brassicaceae family. *Mol. Biol. Evol.* 29, 1241–1254. doi: 10.1093/molbev/msr297
- Cartolano, M., Pieper, B., Lempe, J., Tattersall, A., Huijser, P., Tresch, A., et al. (2015). Heterochrony underpins natural variation in *Cardamine hirsuta* leaf form. *Proc. Natl. Acad. Sci. U.S.A.* 112, 10539–10544. doi: 10.1073/pnas.1419791112
- Chitwood, D. H., Klein, L. L., O'Hanlon, R., Chacko, S., Greg, M., Kitchen, C., et al. (2016). Latent developmental and evolutionary shapes embedded within the grapevine leaf. *New Phytol.* 210, 343–355. doi: 10.1111/nph.13754
- Chuck, G., Cigan, A. M., Saetern, K., and Hake, S. (2007). The heterochronic maize mutant *corngrass1* results from overexpression of a tandem microRNA. *Nat. Genet.* 39, 544–549. doi: 10.1038/ng2001
- Cong, B., Liu, J., and Tanksley, S. D. (2002). Natural alleles at a tomato fruit size quantitative trait locus differ by heterochronic regulatory mutations. *Proc. Natl. Acad. Sci. U.S.A.* 99, 13606–13611. doi: 10.1073/pnas.172520999
- Conway, L. J., and Poethig, R. S. (1997). Mutations of *Arabidopsis thaliana* that transform leaves into cotyledons. *Proc. Natl. Acad. Sci. U.S.A.* 94, 10209–10214. doi: 10.1073/pnas.94.19.10209
- Costa, M. M. R., Yang, S., Critchley, J., Feng, X., Wilson, Y., Langlade, N., et al. (2012). The genetic basis for natural variation in heteroblasty in *Antirrhinum*. *New Phytol.* 196, 1251–1259. doi: 10.1111/j.1469-8137.2012.04347.x
- Cronk, Q. C. B., Needham, I., and Rudall, P. J. (2015). Evolution of catkins: inflorescence morphology of selected Salicaceae in an evolutionary and developmental context. *Front. Plant Sci.* 6:1030. doi: 10.3389/fpls.2015.01030
- Czesnick, H., and Lenhard, M. (2015). Size control in plants — lessons from leaves and flowers. *Cold Spring Harb. Perspect. Biol.* 7:a019190. doi: 10.1101/cshperspect.a019190
- Das Gupta, M., and Nath, U. (2015). Divergence in patterns of leaf growth polarity is associated with the expression divergence of miR396. *Plant Cell* 27, 2785–2799. doi: 10.1105/tpc.15.00196
- De Beer, G. R. (1951). *Embryos and Ancestors*. Oxford: Clarendon Press.
- Donnelly, P. M., Bonetta, D., Tsukaya, H., Dengler, R. E., and Dengler, N. G. (1999). Cell cycling and cell enlargement in developing leaves of *Arabidopsis*. *Dev. Biol.* 215, 407–419. doi: 10.1006/dbio.1999.9443
- Efroni, I., Blum, E., Goldshmidt, A., and Eshed, Y. (2008). A protracted and dynamic maturation schedule underlies *Arabidopsis* leaf development. *Plant Cell* 20, 2293–2306. doi: 10.1105/tpc.107.057521
- Fahlgren, N., Montgomery, T. A., Howell, M. D., Allen, E., Dvorak, S. K., Alexander, A. L., et al. (2006). Regulation of AUXIN RESPONSE FACTOR3 by TAS3 ta-siRNA Affects Developmental Timing and Patterning in *Arabidopsis*. *Curr. Biol.* 16, 939–944. doi: 10.1016/j.cub.2006.03.065
- Fernandez, D. E. (1997). Developmental basis of homeosis in precociously germinating *Brassica napus* embryos: phase change at the shoot apex. *Development* 124, 1149–1157.
- Friedman, W. E. (1999). Expression of the cell cycle in sperm of *Arabidopsis*: implications for understanding patterns of gametogenesis and fertilization in plants and other eukaryotes. *Development* 126, 1065–1075.
- Friedman, W. E., and Carmichael, J. S. (1998). Heterochrony and developmental innovation: evolution of female gametophyte ontogeny in gnetum, a highly Apomorphic seed plant. *Evolution* 52, 1016–1030. doi: 10.1111/j.1558-5646.1998.tb01830.x
- Geuten, K., and Coenen, H. (2013). Heterochronic genes in plant evolution and development. *Front. Plant Sci.* 4:381. doi: 10.3389/fpls.2013.00381
- Gillmor, C. S., Silva-Ortega, C. O., Willmann, M. R., Buendía-Monreal, M., and Poethig, R. S. (2014). The *Arabidopsis* Mediator CDK8 module genes CCT (MED12) and GCT (MED13) are global regulators of developmental phase transitions. *Development* 141, 4580–4589. doi: 10.1242/dev.111229
- Goebel, K. (1908). *Einleitung in die Experimentelle Morphologie der Pflanzen*. Leipzig: B.G. Teubner.
- Gould, S. J. (1977). *Ontogeny and Phylogeny*. Cambridge, MA: Harvard University Press.
- Guo, C., Xu, Y., Shi, M., Lai, Y., Wu, X., Wang, H., et al. (2017). Repression of miR156 by miR159 regulates the timing of the juvenile-to-adult transition in *Arabidopsis*. *Plant Cell* 29, 1293–1304. doi: 10.1105/tpc.16.00975
- Guo, J. (2015). Mechanisms for the evolution of complex and diversely elaborated sepals in Iris identified by comparative analysis of developmental sequences. *Am. J. Bot.* 102, 819–832. doi: 10.3732/ajb.1400519

AUTHOR CONTRIBUTIONS

MB-M conceived and wrote the manuscript. CSG edited and revised the manuscript.

- Haeckel, E. (1875). Die Gastrula und die Eifurchung der Thiere. *Jena. Z. Naturwiss.* 9, 402–508.
- Harrison, C. J. (2017). Development and genetics in the evolution of land plant body plans. *Philos. Trans. R. Soc. B Biol. Sci.* 372:20150490. doi: 10.1098/rstb.2015.0490
- Harrison, C. J., and Morris, J. L. (2018). The origin and early evolution of vascular plant shoots and leaves. *Philos. Trans. R. Soc. B Biol. Sci.* 373:20160496. doi: 10.1098/rstb.2016.0496
- He, J., Xu, M., Willmann, M. R., McCormick, K., Hu, T., Yang, L., et al. (2018). Threshold-dependent repression of SPL gene expression by miR156/miR157 controls vegetative phase change in *Arabidopsis thaliana*. *PLoS Genet.* 14:e1007337. doi: 10.1371/journal.pgen.1007337
- Hudson, C. J., Freeman, J. S., Jones, R. C., Potts, B. M., Wong, M. M. L., Weller, J. L., et al. (2014). Genetic control of heterochrony in *Eucalyptus globulus*. *G3* 4, 1235–1245. doi: 10.1534/g3.114.011916
- Huijser, P., and Schmid, M. (2011). The control of developmental phase transitions in plants. *Development* 138, 4117–4129. doi: 10.1242/dev.063511
- Hunter, C., Sun, H., and Poethig, R. S. (2003). The *Arabidopsis* heterochronic gene ZIPPY Is an ARGONAUTE family member. *Curr. Biol.* 13, 1734–1739. doi: 10.1016/j.cub.2003.09.004
- Hunter, C., Willmann, M. R., Wu, G., Yoshikawa, M., de la Luz Gutiérrez-Nava, M., and Poethig, R. S. (2006). Trans-acting siRNA-mediated repression of ETTIN and ARF4 regulates heteroblasty in *Arabidopsis*. *Development* 133, 2973–2981. doi: 10.1242/dev.02491
- Jia, X. L., Chen, Y. K., Xu, X. Z., Shen, F., Zheng, Q. B., Du, Z., et al. (2017). miR156 switches on vegetative phase change under the regulation of redox signals in apple seedlings. *Sci. Rep.* 7, 1–13. doi: 10.1038/s41598-017-14671-8
- Kenrick, P., and Crane, P. R. (1997). The origin and early evolution of plants on land. *Nature* 389, 33–39. doi: 10.1038/37918
- Keyte, A. L., and Smith, K. K. (2014). Heterochrony and developmental timing mechanisms: changing ontogenies in evolution. *Semin. Cell Dev. Biol.* 34, 99–107. doi: 10.1016/j.semcdb.2014.06.015
- Kim, J. Y., Oh, J. E., Noh, Y. S., and Noh, B. (2015). Epigenetic control of juvenile-to-adult phase transition by the *Arabidopsis* SAGA-like complex. *Plant J.* 83, 537–545. doi: 10.1111/tpj.12908
- Kostyun, J. L., Preston, J. C., and Moyle, L. C. (2017). Heterochronic developmental shifts underlie floral diversity within Jaltomata (Solanaceae). *EvoDevo* 8:17. doi: 10.1186/s13227-017-0080-z
- Kyozuka, J., Tokunaga, H., and Yoshida, A. (2014). Control of grass inflorescence form by the fine-tuning of meristem phase change. *Curr. Opin. Plant Biol.* 17, 110–115. doi: 10.1016/j.pbi.2013.11.010
- Lemmon, Z. H., Park, S. J., Jiang, K., Van Eck, J., Schatz, M. C., and Lippman, Z. B. (2016). The evolution of inflorescence diversity in the nightshades and heterochrony during meristem maturation. *Genome Res.* 26, 1676–1686. doi: 10.1101/gr.207837.116
- Li, J., Wang, Z., Hu, Y., Cao, Y., and Ma, L. (2017). Polycomb group proteins RING1A and RING1B regulate the vegetative phase transition in *Arabidopsis*. *Front. Plant Sci.* 8:867. doi: 10.3389/fpls.2017.00867
- Li, P., and Johnston, M. O. (2000). Heterochrony in plant evolutionary studies through the twentieth century. *Bot. Rev.* 66, 57–88. doi: 10.1007/BF02857782
- Li, S., Liu, L., Zhuang, X., Yu, Y., Liu, X., Cui, X., et al. (2013). MicroRNAs inhibit the translation of target mRNAs on the endoplasmic reticulum in *Arabidopsis*. *Cell* 153, 562–574. doi: 10.1016/j.cell.2013.04.005
- Lippman, Z. B., Cohen, O., Alvarez, J. P., Abu-Abied, M., Pekker, I., Paran, I., et al. (2008). The making of a compound inflorescence in tomato and related nightshades. *PLoS Biol.* 6:e288. doi: 10.1371/journal.pbio.0060288
- Lord, E. M., and Hill, J. P. (1987). “Evidence for heterochrony in the evolution of plant form,” in *Development as an Evolutionary Process*, eds R. A. Raff and E. C. Raff (New York, NY: Alan R. Liss), 47–70.
- Maugarny-Calés, A., and Laufs, P. (2018). Getting leaves into shape: a molecular, cellular, environmental and evolutionary view. *Development* 145:dev161646. doi: 10.1242/dev.161646
- Meinke, D. W. (1992). A homoetic mutant of *Arabidopsis thaliana* with leafy cotyledons. *Science* 258, 1647–1649. doi: 10.1126/science.258.5088.1647
- Naghiloo, S., and Claßen-Bockhoff, R. (2017). Developmental changes in time and space promote evolutionary diversification of flowers: a case study in Dipsacaceae. *Front. Plant Sci.* 8:1665. doi: 10.3389/fpls.2017.01665
- Nikolov, L. A., Tomlinson, P. B., Manickam, S., Endress, P. K., Kramer, E. M., and Davis, C. C. (2014). Holoparasitic Rafflesiaceae possess the most reduced endophytes and yet give rise to the world's largest flowers. *Ann. Bot.* 114, 233–242. doi: 10.1093/aob/mcu114
- Nodine, M. D., and Bartel, D. P. (2010). MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis. *Genes Dev.* 24, 2678–2692. doi: 10.1101/gad.1986710
- Ojeda, D. I., Jaén-Molina, R., Santos-Guerra, A., Caujape-Castells, J., and Cronk, Q. (2017). Temporal, but not spatial, changes in expression patterns of petal identity genes are associated with loss of papillate conical cells and the shift to bird pollination in Macaronesian Lotus (Leguminosae). *Plant Biol.* 19, 420–427. doi: 10.1111/plb.12551
- Park, S. J., Jiang, K., Schatz, M. C., and Lippman, Z. B. (2012). Rate of meristem maturation determines inflorescence architecture in tomato. *Proc. Natl. Acad. Sci. U.S.A.* 109, 639–644. doi: 10.1073/pnas.1114963109
- Peragine, A., Yoshikawa, M., Wu, G., Albrecht, H. L., and Poethig, R. S. (2004). SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting si RNAs in *Arabidopsis*. *Genes Dev.* 18, 2368–2379. doi: 10.1101/gad.1231804.nism
- Pham, T., Day, S. M., Glassford, W. J., Williams, T. M., and Rebeiz, M. (2017). The evolutionary origination of a novel expression pattern through an extreme heterochronic shift. *Evol. Dev.* 19, 43–55. doi: 10.1111/ede.12215
- Pires, N. D., and Dolan, L. (2012). Morphological evolution in land plants: new designs with old genes. *Philos. Trans. R. Soc. B Biol. Sci.* 367, 508–518. doi: 10.1098/rstb.2011.0252
- Poethig, R. S. (2013). *Vegetative Phase Change and Shoot Maturation in Plants*, 1st Edn. New York, NY: Elsevier Inc. doi: 10.1016/B978-0-12-396968-2.00005-1
- Porras, R., and Muñoz, J. M. (2000). Cleistogamous capitulum in *Centaurea* (Asteraceae): heterochronic origin. *Am. J. Bot.* 87, 925–933. doi: 10.2307/2656991
- Pryer, K. M., and Hearn, D. J. (2009). Evolution of leaf form in marsileaceous ferns: evidence for heterochrony. *Evolution* 63, 498–513. doi: 10.1111/j.1558-5646.2008.00562.x
- Ronse de Craene, L. (2018). Understanding the role of floral development in the evolution of angiosperm flowers: clarifications from a historical and physico-dynamic perspective. *J. Plant Res.* 131, 367–393. doi: 10.1007/s10265-018-1021-1
- Rothwell, G. W., Wyatt, S. E., and Tomescu, A. M. F. (2014). Plant evolution at the interface of paleontology and developmental biology: an organism-centered paradigm. *Am. J. Bot.* 101, 899–913. doi: 10.3732/ajb.1300451
- Sablowski, R. (2016). Coordination of plant cell growth and division: collective control or mutual agreement? *Curr. Opin. Plant Biol.* 34, 54–60. doi: 10.1016/j.pbi.2016.09.004
- Sablowski, R., and Carnier Dornelas, M. (2014). Interplay between cell growth and cell cycle in plants. *J. Exp. Bot.* 65, 2703–2714. doi: 10.1093/jxb/ert354
- Sharbel, T. F., Voigt, M. L., Corral, J. M., Galla, G., Kümlehn, J., Klukas, C., et al. (2010). Apomictic and sexual ovules of boechera display heterochronic global gene expression patterns. *Plant Cell* 22, 655–671. doi: 10.1105/tpc.109.072223
- Smith, K. K. (2002). Sequence heterochrony and the evolution of development. *J. Morphol.* 252, 82–97. doi: 10.1002/jmor.10014
- Smith, K. K. (2003). Time's arrow: heterochrony and the evolution of development. *Int. J. Dev. Biol.* 47, 613–621.
- Stone, S. L., Braybrook, S. A., Paula, S. L., Kwong, L. W., Meuser, J., Pelletier, J., et al. (2008). *Arabidopsis* leafy cotyledon2 induces maturation traits and auxin activity: implications for somatic embryogenesis. *Proc. Natl. Acad. Sci. U.S.A.* 105, 3151–3156. doi: 10.1073/pnas.0712364105
- Sun, B., Looi, L. S., Guo, S., He, Z., Gan, E. S., Huang, J., et al. (2014). Timing mechanism dependent on cell division is invoked by Polycomb eviction in plant stem cells. *Science* 343:1248559. doi: 10.1126/science.1248559
- Tian, H. Q., Yuan, T., and Russell, S. D. (2005). Relationship between double fertilization and the cell cycle in male and female gametes of tobacco. *Sex. Plant Reprod.* 17, 243–252. doi: 10.1007/s00497-004-0233-9
- Tomescu, A. M. F., Wyatt, S. E., Hasebe, M., and Rothwell, G. W. (2014). Early evolution of the vascular plant body plan — the missing mechanisms. *Curr. Opin. Plant Biol.* 17, 126–136. doi: 10.1016/j.pbi.2013.11.016
- Tsai, A. Y.-L., and Gazzarrini, S. (2014). Trehalose-6-phosphate and SnRK1 kinases in plant development and signaling: the emerging picture. *Front. Plant Sci.* 5:119. doi: 10.3389/fpls.2014.00119

- Vasconcelos, T. N. C., Lucas, E. J., Faria, J. E. Q., and Prenner, G. (2018). Floral heterochrony promotes flexibility of reproductive strategies in the morphologically homogeneous genus *Eugenia* (Myrtaceae). *Ann. Bot.* 121, 161–174. doi: 10.1093/aob/mcx142
- Wahl, V., Ponnuru, J., Schlereth, A., Arrivault, S., Langenecker, T., Franke, A., et al. (2013). Regulation of flowering by trehalose-6-phosphate signaling in *Arabidopsis thaliana*. *Science* 339, 704–707. doi: 10.1126/science.1230406
- Walcher-Chevillet, C. L., and Kramer, E. M. (2016). Breaking the mold: understanding the evolution and development of lateral organs in diverse plant models. *Curr. Opin. Genet. Dev.* 39, 79–84. doi: 10.1016/j.gde.2016.06.005
- Wang, J., Czech, B., and Weigel, D. (2009). miR156-regulated SPL transcription factors define an endogenous flowering pathway in *Arabidopsis thaliana*. *Cell* 138, 738–749. doi: 10.1016/j.cell.2009.06.014
- Willmann, M. R., Mehalick, A. J., Packer, R. L., and Jenik, P. D. (2011). MicroRNAs regulate the timing of embryo maturation in *Arabidopsis*. *Plant Physiol.* 155, 1871–1884. doi: 10.1104/pp.110.171355
- Wu, G., Park, M. Y., Conway, S. R., Wang, J., Weigel, D., and Poethig, R. S. (2009). The sequential action of miR156 and miR172 regulates developmental timing in *Arabidopsis*. *Cell* 138, 750–759. doi: 10.1016/j.cell.2009.06.031
- Wu, G., and Poethig, R. S. (2006). Temporal regulation of shoot development in *Arabidopsis thaliana* by miR156 and its target SPL3. *Development* 133, 3539–3547. doi: 10.1242/dev.02521
- Xu, M., Hu, T., Smith, M. R., and Poethig, R. S. (2016). Epigenetic regulation of vegetative phase change in *Arabidopsis*. *Plant Cell* 28, 28–41. doi: 10.1105/tpc.15.00854
- Xu, Y., Guo, C., Zhou, B., Li, C., Wang, H., Zheng, B., et al. (2016). Regulation of vegetative phase change by SWI2/SNF2 chromatin remodeling ATPase BRAHMA. *Plant Physiol.* 172, 2416–2428. doi: 10.1104/pp.16.01588
- Xu, M., Leichty, A. R., Hu, T., and Poethig, R. S. (2018). H2A.Z promotes the transcription of MIR156A and MIR156C in *Arabidopsis* by facilitating the deposition of H3K4me3. *Development* 145:dev.152868. doi: 10.1242/dev.152868
- Yamaguchi, N., Huang, J., Xu, Y., Tanoi, K., and Ito, T. (2017). Fine-tuning of auxin homeostasis governs the transition from floral stem cell maintenance to gynoecium formation. *Nat. Commun.* 8:1125. doi: 10.1038/s41467-017-01252-6
- Yang, L., Xu, M., Koo, Y., He, J., and Poethig, R. S. (2013). Sugar promotes vegetative phase change in *Arabidopsis thaliana* by repressing the expression of MIR156A and MIR156C. *eLife* 2:e00260. doi: 10.7554/eLife.00260
- Yu, S., Li, C., Zhou, C. M., Zhang, T. Q., Lian, H., Sun, Y., et al. (2013). Sugar is an endogenous cue for juvenile-to-adult phase transition in plants. *eLife* 2:e00269. doi: 10.7554/eLife.00269
- Yu, S., Zhang, Y., Horrer, D., Zhang, T., Hao, Y., Feng, Y., et al. (2012). Gibberellin regulates the *Arabidopsis* floral transition through miR156-targeted SQUAMOSA promoter binding – like transcription factors. *Plant Cell* 24, 3320–3332. doi: 10.1105/tpc.112.101014

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 © 2018 Buendia-Monreal and Gillmor. 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.



A Phylogenetic Study of the ANT Family Points to a preANT Gene as the Ancestor of Basal and euANT Transcription Factors in Land Plants

Melissa Dipp-Álvarez and Alfredo Cruz-Ramírez*

Molecular and Developmental Complexity Group, Unidad de Genómica Avanzada, Laboratorio Nacional de Genómica para la Biodiversidad, Centro de Investigación y de Estudios Avanzados del IPN, Guanajuato, Mexico

OPEN ACCESS

Edited by:

Verónica S. Di Stilio,
University of Washington,
United States

Reviewed by:

Alejandra Vasco,
National Autonomous University
of Mexico, Mexico
Chi-Lien Cheng,
The University of Iowa, United States

*Correspondence:

Alfredo Cruz-Ramírez
alfredo.cruz@cinvestav.mx

Specialty section:

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

Received: 16 June 2018

Accepted: 08 January 2019

Published: 29 January 2019

Citation:

Dipp-Álvarez M and
Cruz-Ramírez A (2019) A Phylogenetic
Study of the ANT Family Points to a
preANT Gene as the Ancestor
of Basal and euANT Transcription
Factors in Land Plants.
Front. Plant Sci. 10:17.
doi: 10.3389/fpls.2019.00017

Comparative genomics has revealed that members of early divergent lineages of land plants share a set of highly conserved transcription factors (TFs) with flowering plants. While gene copy numbers have expanded through time, it has been predicted that diversification, co-option, and reassembly of gene regulatory networks implicated in development are directly related to morphological innovations that led to more complex land plant bodies. Examples of key networks have been deeply studied in *Arabidopsis thaliana*, such as those involving the AINTEGUMENTA (ANT) gene family that encodes AP2-type TFs. These TFs play significant roles in plant development such as the maintenance of stem cell niches, the correct development of the embryo and the formation of lateral organs, as well as fatty acid metabolism. Previously, it has been hypothesized that the common ancestor of mosses and vascular plants encoded two ANT genes that later diversified in seed plants. However, algae and bryophyte sequences have been underrepresented from such phylogenetic analyses. To understand the evolution of ANT in a complete manner, we performed phylogenetic analyses of ANT protein sequences of representative species from across the Streptophyta clade, including algae, liverworts, and hornworts, previously unrepresented. Moreover, protein domain architecture, selection analyses, and regulatory *cis* elements prediction, allowed us to propose a scenario of how the evolution of ANT genes occurred. In this study we show that a duplication of a preANT-like gene in the ancestor of embryophytes may have given rise to the land plant-exclusive basalANT and euANT lineages. We hypothesize that the absence of euANT-type and basalANT-type sequences in algae, and its presence in extant land plant species, suggests that the divergence of pre-ANT into basal and eu-ANT clades in embryophytes may have influenced the conquest of land by plants, as ANT TFs play important roles in tolerance to desiccation and the establishment, maintenance, and development of complex multicellular structures which either became more complex or appeared in land plants.

Keywords: AP2-like, euANT, basalANT, phylogeny, *cis*-elements

INTRODUCTION

Colonization of Land by Plants

The plant terrestrialization process, also known as the colonization of land by plants, took place 425–490 million years ago (Sanderson, 2003). It has been hypothesized that such process occurred in a microbial moist biofilm as substrate, which could have been composed of fungi, bacteria, cyanobacteria, and green algae (Wellman and Strother, 2015). Prior to land colonization, freshwater green algae related to extant Charophyte algae had already evolved different types of forms; from unicellular organisms to multicellular filaments and thallose forms (Harrison, 2017). It is hypothesized that the first land plants originated from one of these types of ancestral algae from the order Zygnematales that were tolerant to desiccation and adapted to aerial conditions, which allowed them to survive in the previously mentioned substrate (Wellman and Strother, 2015; Ishizaki, 2017; Puttick et al., 2018). In turn, plant terrestrialization was critical for the formation of earth's biosphere, a fundamental change that allowed the colonization of terrestrial environments by animals.

Several innovations allowed early plants to diversify and thrive on changing terrestrial environments, such as the transition from a gametophyte-dominant to sporophyte-dominant life form, the three-dimensional growth form, originated by changes in the division plane orientations of stem cells, the development of a sporophytic apical meristem, and the origin of new organs and tissues such as vasculature, roots, leaves, flowers, and seeds (Pires and Dolan, 2012; Ishizaki, 2017). It has been predicted that such innovations in the land plant body were the product of diversification, co-option, and reassembly of gene regulatory networks implicated in development; given that members of early diverging lineages of land plants share a set of highly conserved transcription factors (TFs) with flowering plants (Pires and Dolan, 2012; Bennett et al., 2014). In line with this idea, a comparative study carried out by Catarino et al. (2016), revealed that the number of TF families has barely increased in land plants after terrestrialization compared to the increase since the divergence of chlorophytes and streptophytes.

From an evolutionary perspective it has been proposed that the AINTEGUMENTA (ANT) family of TFs is conserved as part of the ancestral molecular toolkit that enabled life and development of plants on earth (Kim et al., 2006; Shigyo et al., 2006; Floyd and Bowman, 2007). Previously, it was hypothesized that the common ancestor of mosses and vascular plants encoded two ANT genes that later diversified in seed plants (Floyd and Bowman, 2007). However, algae and bryophyte sequences were underrepresented or lacking from previous phylogenetic analyses (Kim et al., 2006; Shigyo et al., 2006; Floyd and Bowman, 2007) and it is not clear if the origin of the basalANT and euANT subclades predated land plant divergence.

ANT Lineage of Transcription Factors

ANT genes belong to the APETALA 2/ETHYLENE RESPONSE FACTOR (AP2/ERF) family of TFs characterized for containing the AP2/ERF DNA binding domain, which were first discovered

in the *A. thaliana* AP2 protein, a critical protein for the establishment of the floral meristem (Kim et al., 2006). ANT proteins contain two AP2 domains separated by a linker region (25 amino acids) so they fit under the AP2-like subfamily of proteins (Riechmann and Meyerowitz, 1998; **Supplementary Figure 1**). The AP2-like subfamily is further divided in two different lineages: the euAP2 lineage, which mRNA has a miR172 target sequence in the post-domain region, and the AINTEGUMENTA (ANT) lineage, characterized by a 10- amino acid insertion in the first AP2 domain (R1) and a 1- amino acid insertion in the second AP2 domain (R2), see **Supplementary Figure 1**. In turn, the ANT lineage is divided into the basalANT and the euANT sublineages (Horstman et al., 2014). The main difference among eu and basalANT is that euANT proteins are defined by a long pre-domain region and four conserved motifs. The first motif, euANT1, consists of conserved amino acids (NSC[K/R][K/R]EGQ[T/S]R) in the 10-amino acid insertion located in the first AP2 domain. The other three motifs, euANT2 (WLGFSL), euANT3 (PKLEDFLG) and euANT4 (TFGQR), are located in the pre-domain region of the euANT proteins as described in **Supplementary Figure 1** (Kim et al., 2006).

Arabidopsis thaliana euANT TFs have been deeply studied for over 20 years. They are represented in a gene family that include eight members: AINTEGUMENTA (ANT), AINTEGUMENTA-LIKE 1 (AIL1), BABY BOOM (BBM), and PLETHORAS (PLT1, PLT2, PLT3, PLT5, PLT7). They are considered master regulators of diverse developmental processes and are mainly expressed in dividing tissues where they regulate the formation and maintenance of stem cell niches and the correct development of the embryo, the root and shoot organs. The characterization of the *plt1;plt2* double mutant revealed that PLT1 and PLT2 are essential for proper development of *A. thaliana* roots. Both genes are expressed, in an auxin-dependent manner, at the basal part of the octant stage embryo and in the quiescent center in late embryogenesis, and are essential for the correct specification and development of both the embryonic and post-embryonic root meristems (Aida et al., 2004). On the other hand, overexpression (OE) of BBM (PLT4) or PLT5 induces ectopic embryos on the aerial part of the seedling, while PLT1 and PLT2 OE induces ectopic root stem cell niches (reviewed in Horstman et al., 2014).

Arabidopsis thaliana PLT1, PLT2, and PLT3 are expressed in the root forming a gradient distribution with a maximum in the stem cell niche, this is also true for *Oryza sativa* PLT genes (Galinha et al., 2007; Li and Xue, 2011). The gradient distribution of these PLT proteins indicates that they act as a dose-dependent morphogen, because high levels of PLTs promote stem cell identity, while lower levels allow stem cell daughters to partially differentiate and keep mitotic activity (Galinha et al., 2007). However, when cells are displaced to the root region with a drastic lower concentration of PLTs, they lose proliferation capacity and terminally differentiate (Galinha et al., 2007). Seedlings of the homozygous triple mutant *plt1;plt2;plt3* developed by Galinha et al. (2007) show a fully differentiated embryonic root pole at 3 days post-germination (d.p.g.) and they never develop lateral roots. All together, these phenotypes position these specific euANT genes as master regulators of root development. Also, Hofhuis et al. (2013) showed that PLT3, PLT5, and PLT7

are expressed in lateral root primordia, regulating lateral root development and preventing that lateral root primordia are formed close to one another. This study reveals that the genetic mechanism that regulates rhizotaxis and involves *PLT3*, *PLT5*, and *PLT7* is the same that controls phylotaxis in the *A. thaliana* shoot (Prasad et al., 2011).

In the case of *A. thaliana* *ANT*, it has been demonstrated that *ANT* TF plays a key role in flower development (Krizek, 1999). OE of *ANT* under the cauliflower mosaic virus 35S constitutive promoter (35S::*ANT*) causes the development of larger floral organs resulting from an increase in cell number in sepals. While larger petals, stamens, and carpels observed in *ANT* OE plants are the result of an increase in cell size.

Arabidopsis thaliana genes *WRINKLED1* (*WRI1*), *WRINKLED3* (*WRI2*), *WRINKLED4* (*WRI4*), and *ADAP* are members of the basalANT lineage. *WRI1* regulates fatty acid metabolism in the seed, binding to the *cis* region of glycolytic genes and its activity is localized mainly to the maturing embryo (To et al., 2012). *WRI3* and *WRI4* expression is localized to vegetative tissue and flowers where they activate cuticular wax production (Park et al., 2016) and *AthADAP* is necessary for Absciscic acid response and regulates drought response (Lee et al., 2009).

ANT Transcription Factors in Bryophytes

Two basalANT-type TFs, homologs of the rice *SMALL ORGAN SIZE1* (*SMOS1*) gene, were found in *Physcomitrella patens*: *PpSMOS1-like1* and *PpSMOS1-like2*. In rice, *SMOS1* acts downstream of auxins to regulate cell expansion during organ size control and the mutant phenotype, which has an overall smaller size of organs, was partially complemented by *PpSMOS1-like1* (Aya et al., 2014). On the other hand, there are four euANT genes encoded in the *P. patens* genome, *APB1-4*. They act as a molecular switch for the development of different types of stem cells and their expression is induced by auxin, similarly to what has been shown for *A. thaliana* *AIL* genes (Aoyama et al., 2012). No reports exist yet for *ANT* genes in Streptophyte algae and other bryophytes, aside from *Physcomitrella patens*.

ANT Transcription Factors in Non-flowering Vascular Plants

In the C-fern *Ceratopteris richardii*, Bui et al. (2017) identified an ortholog of the *A. thaliana* *ANT* gene, *CrANT*, which has an expression pattern similar to *A. thaliana* *BBM* (in embryo development). The ectopic expression of *CrANT* in *C. richardii* gametophytes originated the spontaneous production of apogamous sporophytes that develop sporophyte-only vascular tissue, tracheids, and stomata. This suggests that *CrANT* has an important role in sexual reproduction and that *AthBBM* might have evolved from an ancestral *AIL* gene such as *CrANT*. A single euANT gene has been reported for both *Gnetum parvifolium* and *Pinus taeda* (gymnosperms) and their expression pattern in young primordia indicate that they might have an important role in lateral organ size control as well as in the development of the ovule (Shigyo and Ito, 2004; Yamada

et al., 2008), very similar to what has been reported for *AthANT* (Klucher et al., 1996; Mizukami and Fischer, 2000).

So far evolutionary analyses that include *ANT* genes from Streptophyte algae and other bryophytes, apart from *P. patens*, have not been reported. Therefore, the main objective of this study was to explore in a more extensive and deep manner the evolutionary relationships among basalANT, euANT and, the first time described, preANT genes. In order to generate this more complete scenario on *ANT* gene evolution along the plant phylogeny, we set the following objectives: (i) to analyze and compare the structures and motif conservation in preANT, basalANT and euANT proteins; (ii) to explore how natural selection has influenced *ANT* gene evolution, (iii) to define differences between *cis*-regulatory elements in the preANT, basalANT, and euANT promoter regions, and (iv) to explore the evolutionary relationships of these genes along the plant kingdom.

METHODS

Phylogenetic Analysis of ANT Proteins

To reconstruct the phylogeny of the *ANT* lineage of TFs among the Streptophyta clade, we conducted a search for basalANT and euANT homologs by querying the coding sequence of eight *A. thaliana* euANT genes (*ANT*, *AIL1*, *PLT1*, *PLT2*, *AIL6*, *BBM*, *PLT5*, and *PLT7*) and three *A. thaliana* basalANT genes (*WRI1*, *ADAP*, and *WRI4*) against Phytozome¹ (Goodstein et al., 2012), the Conifer Genome Integrative Explorer² (Nystedt et al., 2013), the OneKP³ database, the *Klebsormidium nitens* genome project⁴, and the recently published genomes from *Azolla filiculoides* and *Salvinia cucullata* (Li et al., 2018). We then selected the *A. thaliana* AP2 gene and homologous sequences from other seed plants, as outgroups to explore the relationships among *ANT* genes.

Putative *ANT* TF coding sequences from selected angiosperms, gymnosperms, ferns, lycophytes, bryophytes, charophytes, and a chlorophyte, were obtained and translated selecting the longest open reading frame. Our final database included a total of 114 *ANT* protein sequences from 23 different species as well as 10 outgroup sequences (taxa, sequence length, and ID shown on **Supplementary Table 1**). AA sequences were aligned with the MAFFT version 7 online service (Katoh et al., 2017) using the L-INS-i method that incorporates local pairwise alignment information (Katoh et al., 2005). The alignment was manually edited using Jalview 2.10.3 (Waterhouse et al., 2009), deleting uninformative gaps and regions of the alignment that were poorly aligned. The final alignment consists of 124 taxa and contains 1854 alignment positions of which 1748 are phylogenetically informative (**Supplementary Data Sheet 1**). We then identified the best partitioning scheme and the best-fit

¹<https://phytozome.jgi.doe.gov/pz/portal.html>

²ConGenIE.org

³<https://db.cngb.org/onekp/>

⁴http://www.plantmorphogenesis.bio.titech.ac.jp/~algae_genome_project/klebsormidium/

evolution model for each section of the protein alignment (pre-domain, AP2-R1, linker, AP2-R2, and post-domain regions) with PartitionFinder 2.1.1 run on the CIPRES Science Gateway (Miller et al., 2011; Lanfear et al., 2017) with the greedy search algorithm (Lanfear et al., 2012), the RAxML option (Stamatakis, 2014), and the corrected Akaike Information Criterion (AICc) to compare models of molecular evolution. The selected best-fitting models of AA substitution were VT+G+F, for the pre- and post-domain regions; JTTDCMUT+I+G, for the AP2-R1 domain; LG4M+G, for the linker; and JTT+G for the AP2-R2 domain. A Maximum Likelihood (ML) phylogenetic reconstruction was then performed with the whole alignment matrix in RAxML version 8.2.10 (Stamatakis, 2014) on CIPRES with the four distinct models identified by PartitionFinder 2.1.1, the GAMMA model of rate heterogeneity and with 1000 rapid bootstrap inferences to evaluate the analysis. The best tree with bootstrap branch labels was then edited using the iTOL online software (Letunic and Bork, 2016).

ANT Protein Structure Analysis

To characterize the protein structure of ANT genes among Streptophyta, we analyzed the position of the reported ANT-type AP2-R1 and AP2-R2 domains (Riechmann and Meyerowitz, 1998; Kim et al., 2006) in the 99 AA sequences retrieved. The presence of the euANT motifs reported by Kim et al. (2006), as well as new conserved motifs and their AA composition was also assessed using the Multiple Em for Motif Elicitation (MEME) version 4.12.0 online software⁵ (Bailey and Elkan, 1994), run to search for 40 conserved motifs with a minimum and maximum length of 4 and 10 AAs, respectively.

Selection Analysis for ANT Genes

To find signatures of natural selection along the protein coding sequence of euANT and basalANT genes we performed two types of Fixed Effects Likelihood (FEL) analyses (Kosakovsky Pond and Frost, 2005) and a Mixed Effects Model of Evolution (MEME) test (Murrell et al., 2012) with the HyPhy software package in the DataMonkey online server (Kosakovsky Pond and Frost, 2005). The input for both tests was a codon alignment of the protein coding region of all the ANT genes in this study (Supplementary Table 1) and the ML phylogenetic tree deduced with RAxML (see Phylogenetic analysis in Results) (Figure 1). The FEL approaches were designed to measure strength of selection, as well as compare positive and negative pervasive selection on sites in the basalANT clade (basalANT clade as foreground and the rest of the sequences as background) with that in the euANT clade (euANT clade as foreground and the rest of the sequences as background). For each of the FEL tests the ratio of non-synonymous to synonymous substitutions ($w = dN/dS$) was calculated for the foreground and background sets of sequences. Significance was determined for each codon position with a p -value < 0.05 for the test $dN \neq dS$. The MEME approach, in which the w can vary through sites, was carried out to identify episodic diversifying selection affecting a proportion of

branches (Murrell et al., 2012) and significance was assessed at $p < 0.05$.

Putative *cis*-Regulatory Elements in ANT Genes

To identify putative *cis*-regulatory elements (CREs) in the promoter region of ANT genes, a 2.5 kb fragment upstream the ATG of ANT genes from *O. sativa*, *A. thaliana*, *A. trichopoda*, *P. patens*, *M. polymorpha*, and *Chlamydomonas reinhardtii* was extracted using Phytozome⁶. The 2.5 kb 5' region flanking the *K. nitens* ANT gene (KFL_006570020, GenBank:DF237606.1; Hori et al., 2014) was obtained from the National Center for Biotechnology Information (NCBI)⁷ site. Sequences were then analyzed, using the PlantPAN 2.0 Multiple promoter analysis online tool (Chow et al., 2016), for consensus CREs. Results were filtered to keep the number and detailed position of MYB, WUSCHEL-related homeobox (WOX), ethylene response factor (ERF), auxin response factor (ARF), and cytokinin response regulator (ARR-B) *cis*-elements found in each ANT gene promoter region.

RESULTS AND DISCUSSION

Phylogeny of the ANT Lineage in Streptophyta

To better understand the origin and evolution of ANT genes we performed a Maximum Likelihood phylogenetic analysis based on the basalANT and euANT protein sequences we retrieved from representative species across the Streptophyta clade (Supplementary Table 1).

In Figure 1 we show the rooted phylogenetic tree for ANT, which shows that the single *C. reinhardtii* (unicellular Chlorophyte algae) and *Chlorokybus atmophyticus* (unicellular Streptophyte algae) sequences appear as a sister clade to a major clade composed of two clades (53% bootstrap support); the land plant ANT clade and a clade harboring the sequences from *Mesotaenium caldariorum* and *K. nitens*. *Chlamydomonas reinhardtii*, and *C. atmophyticus* sequences have an intra AP2-R1 AA insertion, but this sequence is distinct to that found in euANT proteins, the so called euANT1 motif. On the other hand, sequences from *M. caldariorum* and *K. nitens* already bear the euANT1 motif.

Since this is the first report in which a phylogeny for the ANT group includes Streptophyte algae ANT-like sequences, we hypothesize that *C. reinhardtii* and *C. atmophyticus* ANT-like sequences represent the putative ancestral sequence of the ANT group. Meanwhile, because of the position of the single *M. caldariorum* and *K. nitens* sequences and the presence of the euANT1 motif in the first AP2 domain in these genes, we decided to name these two sequences as preANT (Figure 1, marked with purple). Within the land plant ANT clade two highly supported clades are recovered; the basalANT (98%) (Figure 1, cyan bar) and euANT (71%) (Figure 1, orange bar) clades that

⁵<http://meme-suite.org/>

⁶<https://phytozome.jgi.doe.gov/pz/portal.html>

⁷<https://www.ncbi.nlm.nih.gov/>

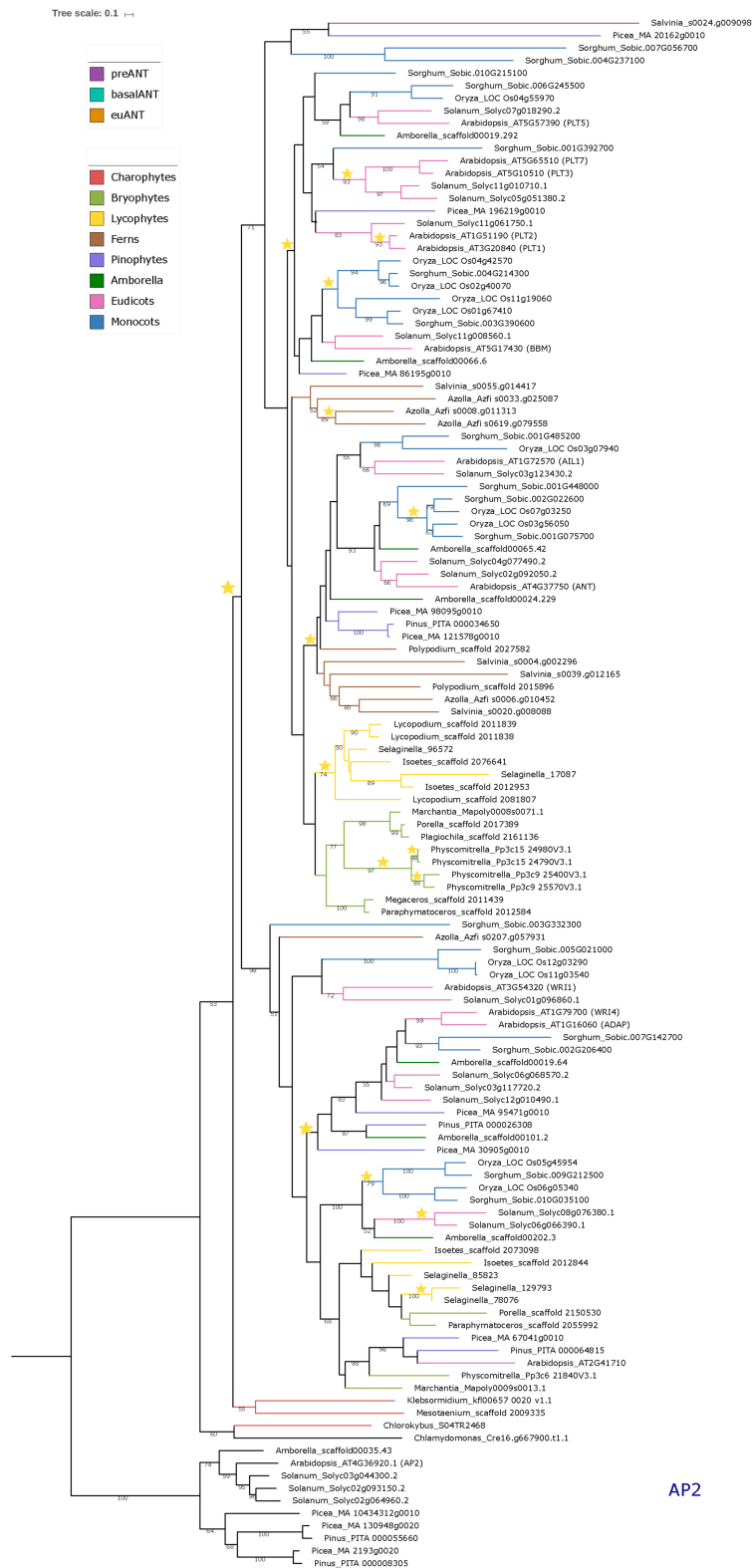


FIGURE 1 | Phylogeny of the ANT gene clade across Streptophyta. Rooted Maximum Likelihood tree constructed on RAxML containing a total of 114 ANT protein sequences from 23 different taxa as well as 10 AP2 outgroup sequences. Bootstrap support was calculated using 1000 replicates. Only bootstrap values over 50% are shown on branches. Colored bar indicates ANT gene clades. Branch colors represent the taxonomic classification to which the taxa belong. Stars represent duplication events. See **Supplementary Table 1** for species names.

were reported in previous phylogenies of ANT genes (Kim et al., 2006; Floyd and Bowman, 2007; Yamada et al., 2008).

The basalANT clade (**Figure 1**, cyan bar) has two subclades; a clade that groups the *AthWRI1* sequence with its closely related sequences from *S. lycopersicum*, *O. sativa* (with two paralogs), and *Sorghum bicolor* (36%), and the other which consists of two subclades: the WRI4/ADAP clade that originated from a duplication event in the ancestor of seed plants (**Figure 1**, yellow star) and that clustered the *AthWRI4* and *AthADAP* sequences together with sequences from *S. lycopersicum*, *S. bicolor*, *A. trichopoda*, *P. taeda*, and *Picea abies* sister to a clade of sequences that includes bryophyte, lycophyte, *Amborella trichopoda*, monocot and eudicot basalANT sequences which clustered with *Arabidopsis*_AT2G41710. All the sequences from the latter clade, share the characteristic of having incomplete AP2-R1 or AP2-R2 domains (see alignment in **Supplementary Figure 2**), which could be the result of different selective pressures driving the fixation of changes that lead to AA losses in the DNA-binding part of the protein in these taxa. Also, we detected duplication events within this clade in lycophytes, and angiosperms (**Figure 1**, yellow stars). WRI1 and WRI4/ADAP clades share a common ancestor with *Azolla*_Azfi_s0207.g057931, a fern sequence (51%). No other fern sequences were nested within the basalANT clade, which could be the result of incomplete sampling or loss of basalANT members.

Little is known about sequences belonging to the basalANT clade, aside from the role of *AthWRI1* in the regulation of fatty acid metabolism in the seed (To et al., 2012), of *AthWRI4* in cuticular wax biosynthesis (Park et al., 2016), and *AthADAP* a key TF for abscisic acid response (Lee et al., 2009). Recent experimental evidence points to the role of ABA in sex determination and transpiration in early land plants, millions of years before the ABA pathway was co-opted to modulate seed dormancy and water balance (McAdam et al., 2016). We can hypothesize that the ancient function of WRI/ADAP proteins could be related mainly to an ABA-dependent pathway that secured adaptation to a desiccated environment via modulating pore function for carbon dioxide and oxygen exchange and controlled water exchange, which requires the generation of an impermeable wax cover. This hypothesis could be tested with the functional characterization of basalANT genes from non-seed plants. We can also speculate that the duplication of WRI/ADAP genes in seed plants (**Figure 1**, yellow star) may have led to the acquisition of new expression patterns by positive selection to regulate fatty acid metabolism, dormancy, and water exchange in the seed and floral tissues.

Two major sister clades formed within the euANT (**Figure 1**, orange bar), one of them includes *AthANT* and *AthAIL1* (48%). All the euANT sequences from bryophytes formed a subgroup within the ANT/AIL1 clade (43%). The sequences from Liverworts (*Porella pinnata*, *Plagiochila asplenoides*, and *M. polymorpha*) clustered together (98%) as a sister clade to the clade containing the four *P. patens* sequences, forming the Setaphyta clade (Puttick et al., 2018). The contrast in number of euANT copies within the moss *P. patens* and the liverworts in this study has to do with the fact that the moss lineage has experienced events of Whole Genome Duplication (WGD) and retention of

paralogs after these events (**Figure 1**, yellow stars) (Lang et al., 2018). The hornwort sequences (*Megaceros tosanus* and *P. hallii*) appeared as a sister clade (100%) to the Liverwort/Mosses group. The sequences from lycophytes (*Isoetes* sp., *Selaginella moellendorffii*, and *Lycopodium deuterodensum*) were grouped together in a clade (74%) sister to the bryophyte sequences. Floyd and Bowman (2007) also recovered the lycophyte sequences in their Bayesian analysis of euANT genes as sister clade to the ANT/AIL1 clade. All the sequences from ferns, were recovered in a sister clade to the bryophyte/lycophyte clade, although with low bootstrap support (38%). This clade also harbors sequences from pinophytes, *A. trichopoda*, eudicots and monocots closely related to *AthANT* and *AthAIL1*. A clade of just *A. filiculoides* and *S. cucullata* sequences was recovered as sister to the major ANT/AIL1 clade, the bootstrap support for this clade is low (<50) but it could have its origin after a duplication of the euANT gene in ferns. It would be interesting to see if this clade maintains its position when more fern genomes become available and a better sampling of ANT genes is possible.

The other euANT major clade in this phylogenetic estimation includes all the PLT/BBM sequences from *A. thaliana* and their closely related sequences from *P. abies*, *A. trichopoda*, *S. lycopersicum*, *O. sativa*, and *S. bicolor*. This clade was also resolved in previous euANT phylogenies (Shigyo and Ito, 2004; Kim et al., 2006; Floyd and Bowman, 2007; Yamada et al., 2008) but no pinophyte sequences were resolved within this clade before. In our phylogeny, *Picea*_MA_86195g0010 was included in the BBM clade, and *Picea*_MA_196219g0010 resolved within the PLT1/2 clade. This, together with the fact that no bryophyte, lycophyte or fern sequences were clustered within this clade, could mean that the PLT/BBM clade had its origins after an event of duplication in the common ancestor of seed plants (**Figure 1**, yellow star). Within the BBM clade, we found a single ortholog in gymnosperms, *Amborella* and eudicots while monocots have two or more copies, originating from a duplication event in the ancestor of monocots and subsequent duplication in *O. sativa*. It is now known that in *O. sativa* four BBM-like genes have a redundant function in early embryogenesis (Khanday et al., 2018), possibly monocots have retained more copies of BBM because of a selection pressure to maintain a redundant function between all the copies in case a mutation arises that could compromise early development of the plant. But as our phylogenetic reconstruction includes species that belong only to the Poaceae family, an extensive sampling of monocots would be helpful to detect if this retention of copies is exclusive to Poaceae or is a trait common to all monocots. The PLT clade resolved in our phylogenetic reconstruction indicates that a possible duplication of an ancestral PLT1-like gene originated the PLT5 and PLT1-2/PLT3-7 clades in angiosperms. None of the euANT sequences from *A. trichopoda* and *O. sativa* were resolved within the PLT1-2/PLT3-7 clade, this could be due to extinction of the members of this clade or to sampling error. The PLT1-2/PLT3-7 clade has diversified more in eudicots than in other angiosperms and there is evidence that paralogs like PLT1-2 and PLT3-7 have redundant function and similar expression patterns (Horstman et al., 2014). We also recovered a small clade (<50% bootstrap support) of sequences belonging to *S. bicolor*, *P. abies*,

and *S. cucullata* sister to the euANT clade, these sequences have more variable residues in the R1-linker-R2 region with respect to the other sequences in the alignment (see alignment matrix in **Supplementary Figure 2**), and could be in the way to pseudogenization via the accumulation of mutations after a possible event of duplication (Moore and Purugganan, 2005).

Overall, considering streptophyte algae and previously underrepresented bryophyte, lycophyte, and fern sequences in the reconstruction of ANT phylogeny helps us broadly understand the evolution of this lineage of TFs, as some nodes within the phylogenetic tree are not significantly supported by resampling. Our results suggest that a duplication of a single preANT ancestor gene originated the land plant-exclusive basalANT and euANT lineages and support that, as Floyd and Bowman (2007) hypothesized, the most recent common ancestor of embryophytes had one basalANT and one euANT protein encoded in its genome. It would be interesting if this hypothesis holds and the relationships between ANT gene clades is resolved with greater support as new genomes of streptophyte algae and non-seed plants are published and added to the phylogenetic analysis of ANT TFs.

ANT Protein Structure Analysis

In order to characterize ANT protein structure among streptophyte organisms, and to identify putative differences that could distinguish groups formed in our phylogenetic analysis, we first focused in analyzing the position and AA composition in AP2-R1 and AP2-R2 domains, based on the previously description of such domains (Riechmann and Meyerowitz, 1998; Kim et al., 2006). For this we used the retrieved sequences (enlisted in **Supplementary Table 1**) and generated a *de novo* motif search using MEME⁴ (Bailey and Elkan, 1994). In the sequences analyzed, euANT proteins range from 263 to 882 AAs in length (**Figure 2A**), while those of basalANT from 309 to 472 AAs (**Figure 2B**). In the proteins we denominated as preANT, lengths are 436, 455, and 803 AAs for *C. atmophyticus*, *M. caldarium*, and *K. nitens*, respectively (**Figure 2B**). All euANT and preANT AA sequences have a long pre-domain region while all basalANT proteins possess a short pre-domain region. This could mean that the ancestral streptophyte ANT protein likely had a long pre-domain region that was lost in the basalANT lineage, contrasting with Kim et al. (2006) that hypothesized that a short pre-domain region was the ancestral state of ANT proteins. Such conclusion may be influenced by the fact that basalANT-like proteins with a short pre-domain were recovered as sister sequences to the basalANT and euANT clades in their phylogenetic reconstruction of ANT while our new phylogenetic reconstruction, which includes sequences from algae, suggests a loss of the long pre-domain region in the basalANT clade.

The respective lengths of AP2-R1 and AP2-R2 domains are highly conserved in the majority of the AA sequences analyzed in this study. The AP2-R1 domain ranges from 72 to 77 AAs and the AP2-R2 domain from 65 to 70 AAs. Some exceptions are the euANT sequence AT5G10510_AthPLT3 that has an AP2-R1 motif of 98 AAs, and AT2G41710_Ath, Solyc08g076380_Sly, and

Os05g45954_Osa all basalANT sequences with expanded AP2-R1 domains of 110, 83, and 89 AAs (**Supplementary Figure 2**). Sequences with a reduced version of the AP2-R1 domain include just euANT sequences from eudicots; Solyc03g123430_Sly, with 64 AAs; Os07g03250_Osa, with 33 AAs, and Os03g07940_Osa with 38 AAs. The fern euANT sequences scaffold_2008839_Aca and scaffold_2027582_Phe possess a reduced version of the AP2-R2 domain, with 17 and 24 AAs, respectively. In the same manner, the *Amborella* sequence scaffold00024.229_Atr has an AP2-R2 domain of 54 AAs and the eudicot sequence Solyc03g123430_Sly, of 60 AAs. Two basalANT sequences also have a different AP2-R2 domain length wise, AT2G41710_Ath and Solyc08g076380_Sly with 15 and 102 AAs, respectively. The length of the AP2 domains in preANT proteins is within the range estimated for the majority of the sequences in this study. It would be interesting to test if ANT proteins with reduced AP2 domains can still bind DNA and, if so, what is the sequence of the *cis*-element to which they bind. We also found that the linker region is deeply conserved between all the sequences analyzed (see also alignment in **Supplementary Figure 2**). It is possible that changes in this region resulted deleterious, this suggests its importance in correct protein structure and function.

Some of the most conserved motifs found by MEME reside within the AP2-R1 – linker – AP2-R2 region of ANT proteins. The euANT1-4 motifs, identified by Kim et al. (2006), were represented in our MEME results (light green boxes in **Figure 2**) in sequences belonging to the euANT clade. One of the euANT sequences from *S. bicolor*, Sobic.007G056700, has a divergent 10-AA insertion in the AP2-R1 domain and therefore does not have the euANT1 motif. Moreover, MEME did not find any other of the euANT motifs in this protein. This divergence is only present in some of the monocot sequences analyzed, as another two sequences from *O. sativa*, Os03g07940 and Os07g03250 appear to have lost all euANT motifs. We found that 69% of euANT sequences have the euANT2 motif, 79% the euANT3 motif and 90% the euANT4 motif (**Figure 2**). A previous euANT lineage characterization stated that all euANT proteins possess the four euANT motifs (Kim et al., 2006) but as more sequences from more taxa are considered, we observe that this is not the rule for all plant lineages.

Other conserved motifs shared by ANT proteins besides euANT1-4 were named M1 through M13. Motifs shared by basalANT and euANT proteins are M5, M7, M11, and M13. The M5 motif (**Figures 2A,B**, cyan boxes) is located downstream of the AP2-R2 domain in basal and euANT sequences. In euANT proteins, M5 is only present in bryophytes and gymnosperms, suggesting that it was lost in ferns and in the lineage that gave rise to angiosperms. M7 motif, present in 53% of the sequences analyzed, is located downstream of the AP2-R2 domain of euANT genes and its sequence suggests it could be a nuclear localization signal (Aida et al., 2004). The M11 motif (**Figures 2A,B**, brown boxes) in the euANT clade is present in lycophytes, gymnosperms, and angiosperms and is absent from bryophytes and ferns, suggesting it appeared in the ancestor of tracheophytes. Only monocot sequences from the basalANT clade share M11, this may be product of a convergent evolution of this motif in both clades or of a loss of this motif in the

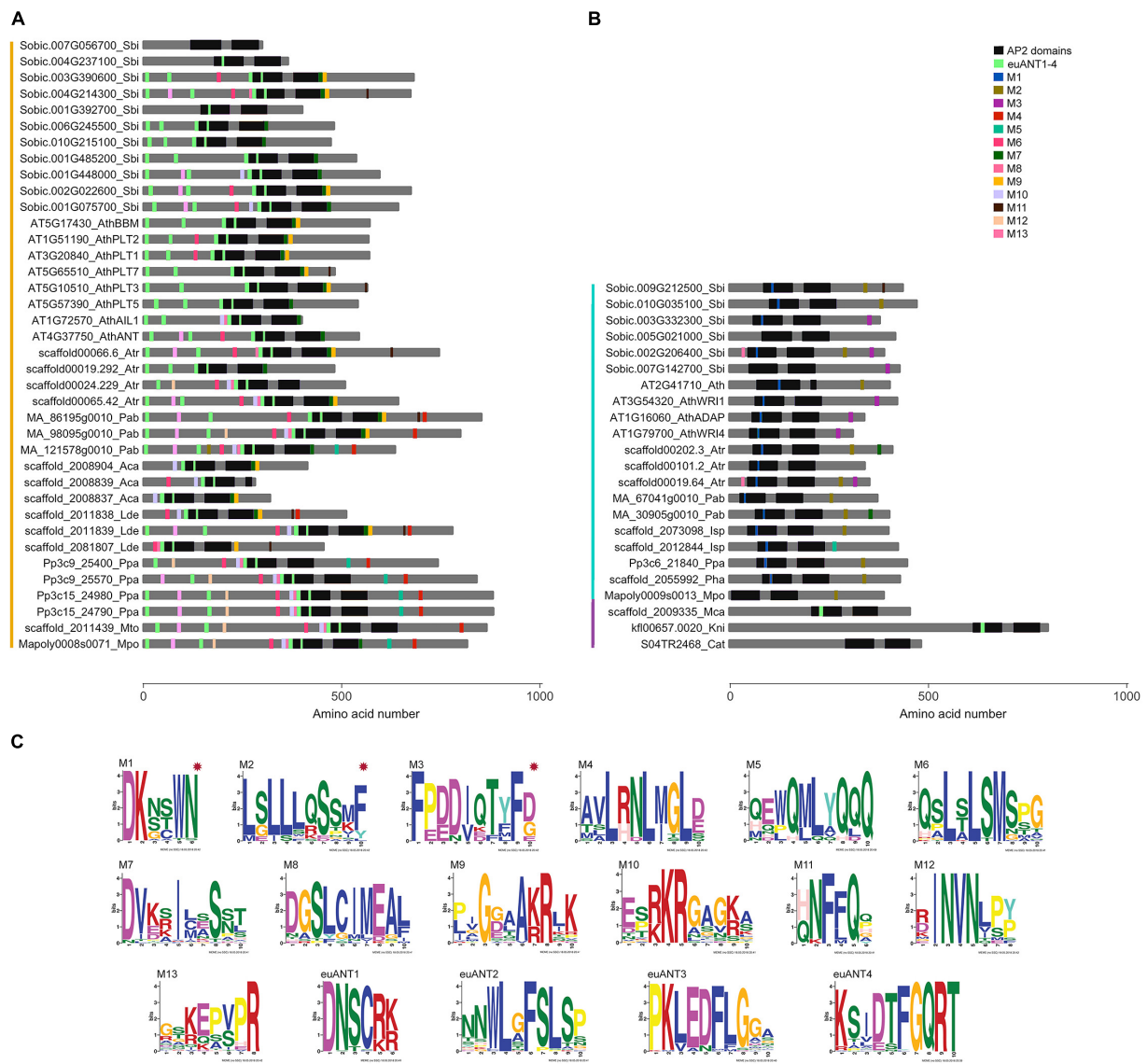


FIGURE 2 | ANT protein conserved motifs in streptophyte sequences. Distribution of motifs identified by MEME in (A) euANT, (B) basalANT, and (C) preANT proteins from representative species of major streptophyte lineages. Each motif (M1 to M13) is represented with a different colored box along a gray line that represents the amino acid (AA) sequence of the protein. Deeply conserved AP2-R1 and AP2-R2 domains are shown as black rectangles of approximately 76 AAs. Motifs euANT1-4, also identified in this study and reported previously by Kim et al. (2006), are displayed with light green boxes. Colored vertical lines left to taxa denote euANT AA sequences in orange, basalANT in cyan and preANT in purple. (C) Logo of each motif identified by MEME. basalANT-exclusive motifs are marked with a red star. See **Supplementary Figure 3** for extended MEME results of all the ANT sequences in this study.

euANT sequences from ferns and basalANT sequences from all lineages except monocots. Also, the M13 motif (**Figures 2A,B**, pink boxes) could have arisen independently in both clades, as euANT proteins belonging to the AIL1/ANT clade (**Figure 1**) and monocot proteins from the BBM clade share it, while only six basalANT proteins from angiosperm species have the M13 motif (**Supplementary Figure 3**).

Motifs exclusive to the basalANT clade of proteins are M1, located within the 10-AA insertion in the AP2-R1 domain (**Figure 2C**, red asterisks); the Leucine-rich M2, situated toward the C-terminus of the protein and shared by 54% of the

sequences and M3. M3 is shared by angiosperm sequences of the WRI/ADAP clades (**Figure 1**). These sequences are the only members of basalANT with known function in fatty acid metabolism and abscisic acid response. It would be interesting to test if the acquisition of the M3 motif in angiosperms has a functional benefit, influencing protein 3D structure, binding to DNA or interaction with other proteins to regulate transcription of target genes.

euANT clade-exclusive motifs are M4, M6, M8, M10, and M12 (**Figure 2C**). M4 is a motif located toward the C-terminal side of the protein and is absent from angiosperm sequences

(Figure 2A, red boxes). M6 is present in sequences from bryophytes, lycophytes, ferns, and gymnosperms, as well as in angiosperm protein sequences that belong to the AIL1/ANT, BBM, and PLT1/PLT2 clades (Figures 1, 2A, in magenta), this suggests M6 was lost before the PLT3/PLT7 – PLT5 divergence. El Ouakfaoui et al. (2010) reported this same motif (M6) in vascular plant sequences and named it euANT5. The M8 motif (Figure 2A) follows the same tendency, with the exception of PLT1/PLT2-type sequences which lacks it. Conserved motifs M10 and M12 (Figure 2A, in light purple and peach boxes) are present in bryophyte, gymnosperm and angiosperm sequences from the AIL1/ANT and BBM clades (Figure 1). We found that the M12 motif was absent from lycophytes and ferns. Our MEME search detected a single conserved motif within the *K. nitens* and *M. caldariorum* sequences, the euANT1 motif. The presence of euANT1 and the length of the pre-domain region in preANT proteins (Figure 2B) suggests that the ancestral ANT sequence resembled more an euANT protein and that subsequent changes in the structure of ANT protein, including the length of the pre-domain region and AP2-R1 domain sequence gave rise to the basalANT lineage of proteins. euANT members from bryophytes possess almost all the euANT conserved motifs found in our MEME analysis, this suggests that the most recent common ancestor of embryophytes already bore these motifs. If the conserved motifs found have a functional role, the loss of motifs in certain angiosperm euANT sequences suggests that subfunctionalization took place after gene duplication and divergence in this lineage.

Whether the differences in motifs among the ANT-like proteins explored in this study, both those first described here and those previously described elsewhere, are determinant for the function of these TFs remains to be explored *in planta* by using the few plant model systems available so far.

Selection Analysis for ANT Genes

The adaptation of plants to a desiccated environment is, in high degree, the result of whole genome duplications and subsequent changes in coding, non-coding and regulatory sequences that augment the diversification potential in plant lineages and that selection will act upon (Van de Peer et al., 2009; Wittkopp and Kalay, 2012). To find evidence of the type and strength of natural selection acting along ANT gene evolution, we performed two types of FEL site-wise selection analyses using the ratio of non-synonymous to synonymous substitutions (dN/dS) along the codon alignment of the euANT and basalANT coding sequences (Kosakovsky Pond and Frost, 2005). When the euANT branches were selected as foreground and basalANT branches as background, the dN/dS value (w), which indicates the ratio of non-synonymous to synonymous substitutions, was 0.3781 for euANT and 0.2720 for basalANT. Similar w values were obtained when basalANT branches were selected as foreground and euANT as background, 0.2704 and 0.3782, respectively. This suggests that the basalANT and euANT clades mainly underwent negative or purifying selection ($w < 1$), which could explain the relatively low sequence variability in the AP2R1-linker-AP2R2 region that binds to DNA and allows transcriptional regulation of ANT target genes. This analysis also suggests

that basalANT genes have been through stronger purifying selection than euANT genes (Supplementary Table 2), which have accumulated more changes in the pre- and post-domain regions after tracheophyte divergence.

The site-specific selection analyses carried out along the 2,612 codons in the ANT genes alignment, revealed that 17 sites are under positive selection and 302 sites are under negative selection constantly along the ANT phylogeny (with a p -value < 0.05 for the test $dN \neq dS$) in the euANT clade as calculated by FEL. While the basalANT clade coding sequences have 6 and 235 sites under constant positive and negative selection pressure, respectively ($p < 0.05$). All the euANT and basalANT sites that are unusually variable and are under positive diversifying selection, fall outside the region that codes for the two AP2 domains and the linker region of the protein (Supplementary Figure 4B). The MEME analysis supported seven of the sites detected by FEL, as having positive selection in the euANT clade and two of the sites in the basalANT clade ($p < 0.05$). MEME test detected that another 34 sites are under episodic diversifying selection ($p < 0.05$, Supplementary Table 2). These results are evidence that the majority of the coding sequence changes, occurring in the pre- and post-domain regions of both basalANT and euANT through time, might have been caused by neutral mutations that did not affect the fitness of the proteins.

Finally, because it has been shown that certain ANT genes respond to the WUSCHEL-related homeobox 5 (WOX5) TF (Ding and Friml, 2010) and to phytohormones like auxins (Aida et al., 2004; Li and Xue, 2011; Aoyama et al., 2012; Yamaguchi et al., 2016), cytokinins (Dello Ioio et al., 2008; Li and Xue, 2011), and ethylene (Li and Xue, 2011), we made an exploratory computational exercise to identify the putative association of diverse transcriptional regulators with the regulatory regions of ANT genes among Streptophyta. For this, we searched for ERE, ARE, ARR-B, and WOX/WUSCHEL-responsive *cis*-elements found in the promoter of ANT genes from *O. sativa*, *A. thaliana*, *A. trichopoda*, *P. patens*, *M. polymorpha*, *K. nitens*, and *C. reinhardtii*.

On average, auxin response elements (AREs) are the most abundant motifs in the ANT regulatory regions analyzed (Supplementary Figure 4A and Supplementary Table 3). euANT genes regulate embryogenesis, stem cell niche maintenance and organ growth both in the shoot and root of *A. thaliana* by their interaction with the phytohormone auxin (Horstman et al., 2014). Also, auxin-induced euANT genes from the moss *P. patens* act as molecular switches for the development of different types of stem cells and three-dimensional growth (Aoyama et al., 2012). Interestingly, it has been reported that the auxin perception pathway composed by TIR1-Aux/IAA-ARF is not present in *K. nitens* (Ohtaka et al., 2017), however, we found four AREs in the promoter region of the preANT gene of this species. In light of these findings, our results suggest that, although AREs are present in the promoter regions of algae ANT genes, their putative ARF-mediated interaction with Auxins was acquired later in the common ancestor of embryophytes and could have allowed the development of 3D apical growth (an innovation in the transition to land), regulation of stem cell identity, and

organ size control (Harrison, 2017). The second most abundant CRE is the WOX/WUSCHEL TF family DNA-binding motif (Supplementary Figure 4). All the sequences analyzed contain at least one WOX/WUSCHEL-responsive motif, being potential targets of such TFs, except for the promoter region of the basalANT gene *Mapoly0009s0013* (Supplementary Table 3). It has been demonstrated that WOX5 acts downstream of Auxin to regulate PLT1 expression in the root apical meristem of *A. thaliana* (Ding and Friml, 2010). Green algae and basal plant lineages contain WOX genes belonging to the ancient clade of WOX TFs and possess the conserved homeodomain motif that binds to DNA to regulate transcription of target genes (Lian et al., 2014). From our *in silico* exploration, we hypothesize that a WOX-ANT genetic interaction could have arisen in the common ancestor of streptophytes. We speculate that this ancient interaction, in which a WOX TF acts upstream regulating the expression of an ANT gene, may have been conserved in modern plants. A possible example of this is the WOX5- *PLT1* genetic interaction, two major factors for root stem cell niche maintenance in *A. thaliana* (Aida et al., 2004). Potential transcriptional regulation of ANT genes in response to cytokinin can also be inferred by the presence of ARR-B CREs in promoter sequences representative of every major streptophyte lineage (Supplementary Table 3). Holzinger and Becker (2015) found that orthologs of the cytokinin signaling in the streptophyte algae *Klebsormidium cernulatum* were up-regulated upon desiccation. Although the interaction between cytokinin and ANT genes could have been originated in the common ancestor of streptophytes, it would be interesting to test experimentally if the angiosperm basalANT genes *WR11*, *WR14*, and *ADAP* are direct transcriptional targets of ARR-B because of their role in adaptation to desiccation. We also found that on average, ERF-responsive CREs (named GCC boxes; Fujimoto et al., 2000) were the least common *cis*-elements in ANT genes from streptophytes. Only four basalANT promoters had GCC boxes and corresponded to angiosperm sequences (Supplementary Table 3), this suggests that a putative interaction between ERFs and basalANT genes was acquired after the duplication of this genes in angiosperms.

Our results on the presence and abundance of these CREs in the promoter regions of the analyzed pre, basal and euANT genes,

are only predictive for those genes and species. So far, there is no experimental evidence that indicates they do respond to a given hormone or that TFs indeed bind to them. To demonstrate that a specific hormone influences, through the binding of a given TF, gene transcription future wet-lab experiments are needed.

In summary, here we explored and analyzed in a more complete and deep manner the putative origin of basalANT and euANT TFs. This is the first study where streptophyte algae sequences have been identified and analyzed for these TFs. Moreover, our approach also explores differences in protein motifs and *cis*-regulatory elements in the regulatory sequences of representative genes along the plant tree of life. Finally, our natural selection analysis reveals those sequences of the genes that have been subject to changes versus those which remain highly conserved. Altogether, our results provide a broad framework for ANT-like genes putative functions in the evolution of plant development, which we assume is useful for the generation of novel hypotheses about protein motif importance and *cis*-elements relevance in ANT gene expression that could be validated experimentally, in order to define the relevance of ANT TFs in the conquest of land by plants.

AUTHOR CONTRIBUTIONS

MD-Á and AC-R conceived the study and wrote the manuscript. MD-Á performed analyses.

ACKNOWLEDGMENTS

MD-Á wishes to thank to Consejo Nacional de Ciencia y Tecnología (CONACYT) for the scholarship Grant No. 487660. AC-R wish to thank CINVESTAV for Research Annual Budget.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.00017/full#supplementary-material>

REFERENCES

- Aida, M., Beis, D., Heidstra, R., Willemsen, V., Blilou, I., Galinha, C., et al. (2004). The PLETHORA genes mediate patterning of the Arabidopsis root stem cell niche. *Cell* 119, 109–120. doi: 10.1016/j.cell.2004.09.018
- Aoyama, T., Hiwatashi, Y., Shigyo, M., Kofuji, R., Kubo, M., Ito, M., et al. (2012). AP2-type transcription factors determine stem cell identity in the moss *Physcomitrella patens*. *Development* 139, 3120–3129. doi: 10.1242/dev.076091
- Aya, K., Hobo, T., Sato-Izawa, K., Ueguchi-Tanaka, M., Kitano, H., and Matsuoka, M. (2014). A novel AP2-type transcription factor, small organ SIZE1, controls organ size downstream of an auxin signaling pathway. *Plant Cell Physiol.* 55, 897–912. doi: 10.1093/pcp/pcu023
- Bailey, T. L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in bipolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 2, 28–36.
- Bennett, T. A., Liu, M. M., Aoyama, T., Bierfreund, N. M., Braun, M., Coudert, Y., et al. (2014). Plasma membrane-targeted PIN proteins drive shoot development in a moss. *Curr. Biol.* 24, 2776–2785. doi: 10.1016/j.cub.2014.09.054
- Bui, L. T., Pandzic, D., Youngstrom, C. E., Wallace, S., Irish, E. E., Szövényi, P., et al. (2017). A fern AINTEGUMENTA gene mirrors baby boom in promoting apogamy in *Ceratopteris richardii*. *Plant J.* 90, 122–132. doi: 10.1111/tpj.13479
- Catarino, B., Hetherington, A. J., Emms, D. M., Kelly, S., and Dolan, L. (2016). The stepwise increase in the number of transcription factor families in the precambrian predated the diversification of plants on land. *Mol. Biol. Evol.* 33, 2815–2819. doi: 10.1093/molbev/msw155
- Chow, C. N., Zheng, H. Q., Wu, N. Y., Chien, C. H., Huang, H., Da Lee, T. Y., et al. (2016). PlantPAN 2.0: an update of plant promoter analysis navigator for reconstructing transcriptional regulatory networks in plants. *Nucleic Acids Res.* 44, D1154–D1164. doi: 10.1093/nar/gkv1035
- Dello Ioio, R., Nakamura, K., Moubayidin, L., Perilli, S., Taniguchi, M., Morita, M. T., et al. (2008). A genetic framework for the control of cell division and

- differentiation in the root meristem. *Science* 322, 1380–1384. doi: 10.1126/science.1164147
- Ding, Z., and Friml, J. (2010). Auxin regulates distal stem cell differentiation in Arabidopsis roots. *Proc. Natl. Acad. Sci. U.S.A.* 107, 12046–12051. doi: 10.1073/pnas.1000672107
- El Ouakfaoui, S., Schnell, J., Abdeen, A., Colville, A., Labbé, H., Han, S., et al. (2010). Control of somatic embryogenesis and embryo development by AP2 transcription factors. *Plant Mol. Biol.* 74, 313–326. doi: 10.1007/s11103-010-9674-8
- Floyd, S. K., and Bowman, J. L. (2007). The ancestral developmental tool kit of land plants. *Int. J. Plant Sci.* 168, 1–35. doi: 10.1086/509079
- Fujimoto, S. Y., Ohta, M., Usui, A., Shinshi, H., and Ohme-Takagi, M. (2000). Arabidopsis ethylene-responsive element binding factors act as transcriptional activators or repressors of gcc box-mediated gene expression. *Plant Cell* 12, 393–404. doi: 10.1105/tpc.12.3.393
- Galinha, C., Hofhuis, H., Luijten, M., Willemsen, V., Blilou, I., Heidstra, R., et al. (2007). PLETHORA proteins as dose-dependent master regulators of Arabidopsis root development. *Nature* 449, 1053–1057. doi: 10.1038/nature06206
- Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, D1178–D1186. doi: 10.1093/nar/gkr944
- Harrison, C. J. (2017). Development and genetics in the evolution of land plant body plans. *Philos. Trans. R. Soc. B* 372, 20150490. doi: 10.1098/rstb.2015.0490
- Hofhuis, H., Laskowski, M., Du, Y., Prasad, K., Grigg, S., Pinon, V., et al. (2013). Phyllotaxis and rhizotaxis in Arabidopsis are modified by three plethora transcription factors. *Curr. Biol.* 23, 956–962. doi: 10.1016/j.cub.2013.04.048
- Holzinger, A., and Becker, B. (2015). Desiccation tolerance in the streptophyte green alga *Klebsormidium*: the role of phytohormones. *Commun. Integr. Biol.* 8:e1059978. doi: 10.1080/19420889.2015.1059978
- Hori, K., Maruyama, F., Fujisawa, T., Togashi, T., Yamamoto, N., Seo, M., et al. (2014). *Klebsormidium flaccidum* genome reveals primary factors for plant terrestrial adaptation. *Nat. Commun.* 5:3978. doi: 10.1038/ncomms4978
- Horstman, A., Willemsen, V., Boutilier, K., and Heidstra, R. (2014). Aintegumenta-like proteins: hubs in a plethora of networks. *Trends Plant Sci.* 19, 146–157. doi: 10.1016/j.tplants.2013.10.010
- Ishizaki, K. (2017). Evolution of land plants: insights from molecular studies on basal lineages. *Biosci. Biotechnol. Biochem.* 81, 73–80. doi: 10.1080/09168451.2016.1224641
- Katoh, K., Kuma, K. I., Toh, H., and Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* 33, 511–518. doi: 10.1093/nar/gki198
- Katoh, K., Rozewicki, J., and Yamada, K. D. (2017). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* doi: 10.1093/bib/bbx108 [Epub ahead of print].
- Khanday, I., Skinner, D., Yang, B., Mercier, R., and Sundaresan, V. (2018). A male-expressed rice embryogenic trigger redirected for asexual propagation through seeds. *Nature* 565, 91–95. doi: 10.1038/s41586-018-0785-8
- Kim, S., Soltis, P. S., Wall, K., and Soltis, D. E. (2006). Phylogeny and domain evolution in the APETALA2-like gene family. *Mol. Biol. Evol.* 23, 107–120. doi: 10.1093/molbev/msj014
- Klucher, K. M., Chow, H., Reiser, L., and Fischer, R. L. (1996). The aintegumenta gene of arabidopsis required for ovule and female gametophyte development is related to the floral homeotic gene APETALA2. *Plant Cell Online* 8, 137–153. doi: 10.1105/tpc.8.2.137
- Kosakovsky Pond, S. L., and Frost, S. D. W. (2005). Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.* 22, 1208–1222. doi: 10.1093/molbev/msi105
- Krizek, B. A. (1999). Ectopic expression of Aintegumenta in Arabidopsis plants results in increased growth of floral organs. *Dev. Genet.* 25, 224–236. doi: 10.1002/(SICI)1520-6408(1999)25:3<224::AID-DVG5>3.0.CO;2-Y
- Lanfear, R., Calcott, B., Ho, S. Y. W., and Guindon, S. (2012). PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol. Biol. Evol.* 29, 1695–1701. doi: 10.1093/molbev/mss020
- Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., and Calcott, B. (2017). Partitionfinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* 34, 772–773. doi: 10.1093/molbev/msw260
- Lang, D., Ullrich, K. K., Murat, F., Fuchs, J., Jenkins, J., Haas, F. B., et al. (2018). The *Physcomitrella patens* chromosome-scale assembly reveals moss genome structure and evolution. *Plant J.* 93, 515–533. doi: 10.1111/tpj.13801
- Lee, S.-J., Cho, D.-I., Kang, J.-Y., and Kim, S. Y. (2009). An ARIA-interacting AP2 domain protein is a novel component of ABA signaling. *Mol. Cells* 27, 409–416. doi: 10.1007/s10059-009-0058-3
- Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44, W242–W245. doi: 10.1093/nar/gkw290
- Li, F.-W., Brouwer, P., Carretero-Paulet, L., Cheng, S., de Vries, J., Delaux, P.-M., et al. (2018). Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nat. Plants* 4, 460–472. doi: 10.1038/s41477-018-0188-8
- Li, P., and Xue, H. (2011). Structural characterization and expression pattern analysis of the rice PLT gene family. *Acta Biochim. Biophys. Hung.* 43, 688–697. doi: 10.1093/abbs/gmr068.Advance
- Lian, G., Ding, Z., Wang, Q., Zhang, D., and Xu, J. (2014). Origins and evolution of WUSCHEL-related homeobox protein family in plant kingdom. *Sci. World J.* 2014, 1–12. doi: 10.1155/2014/534140
- McAdam, S. A. M., Brodribb, T. J., Banks, J. A., Hedrich, R., Atallah, N. M., Cai, C., et al. (2016). Abscissic acid controlled sex before transpiration in vascular plants. *Proc. Natl. Acad. Sci. U.S.A.* 113, 12862–12867. doi: 10.1073/pnas.1606614113
- Miller, M. A., Pfeiffer, W., and Schwartz, T. (2011). “The CIPRES science gateway,” in *Proceedings of the 2011 TeraGrid Conference on Extreme Digital Discovery - TG '11*, New York, NY: ACM Press. doi: 10.1145/2016741.2016785
- Mizukami, Y., and Fischer, R. L. (2000). Plant organ size control: aintegumenta regulates growth and cell numbers during organogenesis. *Proc. Natl. Acad. Sci. U.S.A.* 97, 942–947. doi: 10.1073/pnas.97.2.942
- Moore, R. C., and Purugganan, M. D. (2005). The evolutionary dynamics of plant duplicate genes. *Curr. Opin. Plant Biol.* 8, 122–128. doi: 10.1016/j.pbi.2004.12.001
- Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., and Kosakovsky Pond, S. L. (2012). Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* 8:e1002764. doi: 10.1371/journal.pgen.1002764
- Nystedt, B., Street, N. R., Wetterbom, A., Zuccolo, A., Lin, Y. C., Scofield, D. G., et al. (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature* 497, 579–584. doi: 10.1038/nature12211
- Ohtaka, K., Hori, K., Kanno, Y., Seo, M., and Ohta, H. (2017). Primitive auxin response without TIR1 and Aux/IAA in the charophyte alga *Klebsormidium nitens*. *Plant Physiol.* 174, 1621–1632. doi: 10.1104/pp.17.00274
- Park, C. S., Go, Y. S., and Suh, M. C. (2016). Cuticular wax biosynthesis is positively regulated by WRINKLED4, an AP2/ERF-type transcription factor, in Arabidopsis stems. *Plant J.* 88, 257–270. doi: 10.1111/tpj.13248
- Pires, N. D., and Dolan, L. (2012). Morphological evolution in land plants: new designs with old genes. *Philos. Trans. R. Soc. B* 367, 508–518. doi: 10.1098/rstb.2011.0252
- Prasad, K., Grigg, S. P., Barkoulas, M., Yadav, R. K., Sanchez-Perez, G. F., Pinon, V., et al. (2011). Arabidopsis PLETHORA transcription factors control phyllotaxis. *Curr. Biol.* 21, 1123–1128. doi: 10.1016/j.cub.2011.05.009
- 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.e2–745.e2. doi: 10.1016/j.cub.2018.01.063
- Riechmann, J. L., and Meyerowitz, E. M. (1998). The AP2/EREBP family of plant transcription factors. *Biol. Chem.* 379, 633–646. doi: 10.1515/bchm.1998.379.6.633
- Sanderson, M. J. (2003). Molecular data from 27 proteins do not support a Precambrian origin of land plants. *Am. J. Bot.* 90, 954–956. doi: 10.3732/ajb.90.6.954
- Shigyo, M., Hasebe, M., and Ito, M. (2006). Molecular evolution of the AP2 subfamily. *Gene* 366, 256–265. doi: 10.1016/j.gene.2005.08.009
- Shigyo, M., and Ito, M. (2004). Analysis of gymnosperm two-AP2-domain-containing genes. *Dev. Genes Evol.* 214, 105–114. doi: 10.1007/s00427-004-0385-5

- Stamatakis, A. (2014). RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- To, A., Joubes, J., Barthole, G., Lecureuil, A., Scagnelli, A., Jasinski, S., et al. (2012). WRINKLED transcription factors orchestrate tissue-specific regulation of fatty acid biosynthesis in arabidopsis. *Plant Cell* 24, 5007–5023. doi: 10.1105/tpc.112.106120
- Van de Peer, Y., Maere, S., and Meyer, A. (2009). The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* 10, 725–732. doi: 10.1038/nrg2600
- Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191. doi: 10.1093/bioinformatics/btp033
- Wellman, C. H., and Strother, P. K. (2015). The terrestrial biota prior to the origin of land plants (embryophytes): a review of the evidence. *Palaeontology* 58, 601–627. doi: 10.1111/pala.12172
- Wittkopp, P. J., and Kalay, G. (2012). Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat. Rev. Genet.* 13, 59–69. doi: 10.1038/nrg3095
- Yamada, T., Hirayama, Y., Imaichi, R., and Kato, M. (2008). Aintegumenta homolog expression in *Gnetum* (gymnosperms) and implications for the evolution of ovulate axes in seed plants. *Evol. Dev.* 10, 280–287. doi: 10.1111/j.1525-142X.2008.00237.x
- Yamaguchi, N., Jeong, C. W., Nole-Wilson, S., Krizek, B. A., and Wagner, D. (2016). Aintegumenta and Aintegumenta-Like6/PLETHORA3 induce leafy expression in response to auxin to promote the onset of flower formation in arabidopsis. *Plant Physiol.* 170, 283–293. doi: 10.1104/pp.15.00969

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 Dipp-Álvarez and Cruz-Ramírez. 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.



Co-expression and Transcriptome Analysis of *Marchantia polymorpha* Transcription Factors Supports Class C ARFs as Independent Actors of an Ancient Auxin Regulatory Module

Eduardo Flores-Sandoval¹, Facundo Romani² and John L. Bowman^{1*}

¹ School of Biological Sciences, Monash University, Melbourne, VIC, Australia, ² Facultad de Bioquímica y Ciencias Biológicas, Centro Científico Tecnológico CONICET Santa Fe, Instituto de Agrobiotecnología del Litoral, Universidad Nacional del Litoral – CONICET, Santa Fe, Argentina

OPEN ACCESS

Edited by:

Verónica S. Di Stilio,
University of Washington,
United States

Reviewed by:

Bharti Sharma,
California State Polytechnic University,
Pomona, United States
Raffaele Dello Iorio,
Università degli Studi di Roma La
Sapienza, Italy

*Correspondence:

John L. Bowman
john.bowman@monash.edu

Specialty section:

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

Received: 15 June 2018

Accepted: 27 August 2018

Published: 01 October 2018

Citation:

Flores-Sandoval E, Romani F and
Bowman JL (2018) Co-expression
and Transcriptome Analysis
of *Marchantia polymorpha*
Transcription Factors Supports
Class C ARFs as Independent Actors
of an Ancient Auxin Regulatory
Module. *Front. Plant Sci.* 9:1345.
doi: 10.3389/fpls.2018.01345

We performed differential gene expression (DGE) and co-expression analyses with genes encoding components of hormonal signaling pathways and the ~400 annotated transcription factors (TFs) of *M. polymorpha* across multiple developmental stages of the life cycle. We identify a putative auxin-related co-expression module that has significant overlap with transcripts induced in auxin-treated tissues. Consistent with phylogenetic and functional studies, the class C ARF, MpARF3, is not part of the auxin-related co-expression module and instead is associated with transcripts enriched in gamete-producing gametangiophores. We analyze the MpARF3 and MpmiR160 mutant transcriptomes in the context of coexpression to suggest that MpARF3 may antagonize the reproductive transition via activating the MpMIR11671 and MpMIR529c precursors whose encoded microRNAs target *SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE* (SPL) transcripts of MpSPL1 and MpSPL2. Both MpSPL genes are part of the MpARF3 co-expression group corroborating their functional significance. We provide evidence of the independence of MpARF3 from the auxin-signaling module and provide new testable hypotheses on the role of auxin-related genes in patterning meristems and differentiation events in liverworts.

Keywords: auxin, co-expression, *Marchantia*, auxin response factors, reproductive transitions, class C ARFs, transcriptome, miR160

INTRODUCTION

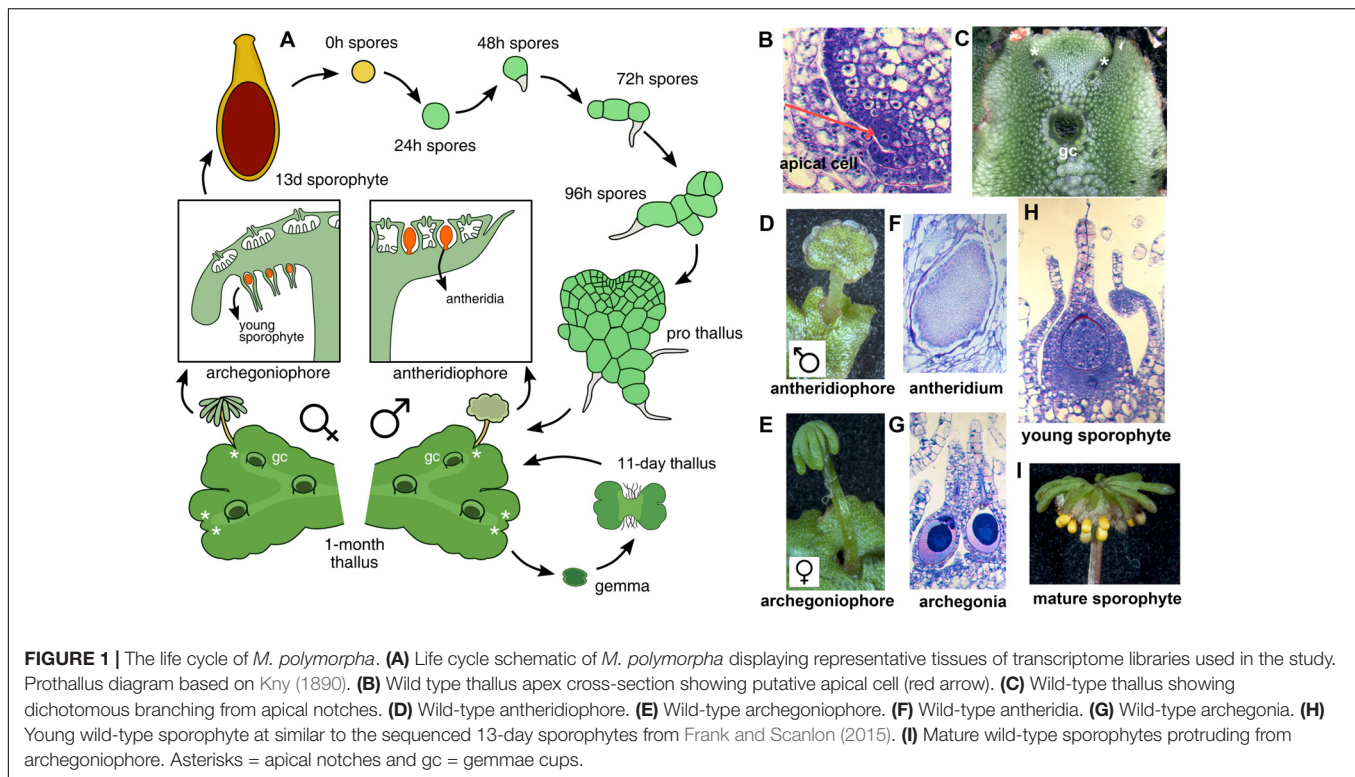
Transcriptome studies in model systems provide a foundation to characterize genetic pathways in the absence of mutational studies. Transcriptome co-expression studies can highlight how specification of complex phenotypes depends on the activity of coordinated batteries of regulatory genes. In plants, many co-expression analyses focused on secondary metabolite production (Ruprecht et al., 2011; Ruprecht and Persson, 2012; Cassan-Wang et al., 2013; Ferreira et al., 2016; Turco et al., 2017), although co-expressed modules for hormonal genes (Ruprecht et al., 2017b; Singh et al., 2017), or developmental factors have also been described (Ichihashi et al., 2014). The functional significance of co-expression modules has been tested by further differential gene

expression (DGE), protein-protein interaction (Kemmeren et al., 2002; Ichihashi et al., 2014), epigenetic (Turco et al., 2017) or functional analyses (Yokoyama et al., 2007; Ruprecht et al., 2011; Cassan-Wang et al., 2013; Ichihashi et al., 2014). Furthermore, studies have corroborated conservation of sets of co-expressed genes in closely related species (Ficklin and Feltus, 2011; Ichihashi et al., 2014; Netotea et al., 2014; Ferreira et al., 2016) as well as in deep evolutionary time (Stuart et al., 2003; Gerstein et al., 2014; Ruprecht et al., 2017a). From a simple ancestral genetic network, gene/genome duplication events could trigger the formation of novel co-expression clusters that can be re-wired and co-opted to pattern novel biological processes (Ruprecht et al., 2017b).

With the establishment of model bryophytes, an increasing number of transcriptome datasets describing multiple developmental and environmental conditions in mosses (Devos et al., 2016; Ortiz-Ramírez et al., 2016; Perroud et al., 2018) and liverworts (Alaba et al., 2015; Frank and Scanlon, 2015; Higo et al., 2016), are providing novel avenues to perform reverse genetic studies in non-vascular plant lineages. *Marchantia polymorpha* is a dioecious liverwort model system (Ishizaki et al., 2016) amenable to gene editing (Sugano et al., 2014) and gene silencing (Flores-Sandoval et al., 2016) techniques. Regulatory genes encoded in the *M. polymorpha* genome exhibit little genetic redundancy, with many transcription factor (TF) families represented by single paralogs (Bowman et al., 2017), making it an attractive system to characterize gene families that display genetic redundancy in other systems. The life cycle of *M. polymorpha* is haploid dominant (**Figure 1A**), with the diploid generation dependent upon the maternal haploid plant. Haploid spores germinate producing a sporeling, a developmental stage with active cell proliferation that forms a simple polarity between anchoring rhizoids and photosynthetic cells (Shimamura, 2016). In accordance with previous work (Bowman et al., 2017; Boehm et al., 2017), day 1 sporelings are single celled but actively differentiate chloroplasts, day 2 sporelings initiate a rhizoid cell, in day 3 sporelings the rhizoid cell elongates and the first divisions of the shoot occur, clearly establishing anatomical apical-basal polarity. By day 4, sporelings undergo proliferation of photosynthetic shoot tissue and further elongation of the rhizoid (**Figure 1A**). Using light cues, sporelings transition into prothalli, a two-dimensional heart-shaped body in which a meristematic zone with an apical cell producing derivatives in two planes is established (Flores-Sandoval et al., 2015). When the apical cell (**Figure 1B**) transitions to producing derivatives in four planes (top, bottom and lateral planes) a three-dimensional vegetative thallus (**Figure 1C**) is formed. Growth in thalli occurs at apical regions with differentiated tissues produced from apical cell derivatives and the apical meristem dichotomously branching (**Figure 1C**) at regular intervals (Shimamura, 2016; Solly et al., 2017). Upon inductive far-red to red light ratios (Kubota et al., 2014; Inoue et al., 2016; Linde et al., 2017) apices differentiate umbrella-like sexual gametangiophores (**Figures 1D,E**) wherein gametes (**Figures 1F,G**) are produced (Higo et al., 2016; Koi et al., 2016; Rövekamp et al., 2016; Yamaoka et al., 2018). In an moist environment, a sperm cell swims and fertilizes an egg cell, thus forming a zygote that undergoes several rounds of cell proliferation (**Figure 1H**) before

differentiating sporogenous (**Figure 1I**) tissue (Shimamura, 2016).

Auxin is a key phytohormone that triggers multiple aspects of land plant physiology and development in a context dependent fashion. Comparative analysis between charophycean algae and land plant genomes and transcriptomes indicates that the canonical F-box mediated auxin signaling pathway evolved in the land plant ancestor (Bowman et al., 2017). Surprisingly, charophycean algae are able to transcriptionally respond to exogenous auxin (Hori et al., 2014; Ohtaka et al., 2017; Romani, 2017; Mutte et al., 2018) even in the absence of specific TIR1 F-box receptors, AUX/IAAs with auxin-sensitive degron (II) domains or distinct class A and B ARFs (Bowman et al., 2017; Flores-Sandoval et al., 2018; Mutte et al., 2018). *M. polymorpha* possesses single paralogs of all the canonical auxin-signaling pathway genes, retaining the predicted ancestral embryophyte state (Flores-Sandoval et al., 2015; Kato et al., 2015; Bowman et al., 2017; Mutte et al., 2018). In *M. polymorpha*, the class A ARF (MpARF1) is necessary to elicit physiological and transcriptional responses to exogenous auxin, acting as a transcriptional activator (Kato et al., 2017). The class B ARF (MpARF2) has not yet been characterized but is regarded as a transcriptional repressor (Kato et al., 2015). The class C ARF (MpARF3) is a target of the embryophyte-specific *miRNA160* and antagonizes differentiation events in multiple developmental contexts with loss-of-function *Mparf3* alleles able to respond to exogenous auxin (Flores-Sandoval et al., 2018; Mutte et al., 2018), and strong gain-of-function MpARF3 alleles resembling auxin-depleted mutants (Eklund et al., 2015; Flores-Sandoval et al., 2018). Phylogenetic analysis indicates Class C ARFs evolved in a charophycean algal ancestor prior to the emergence of the auxin-signaling module (Flores-Sandoval et al., 2018; Mutte et al., 2018). Two additional genes, the NON-CANONICAL ARF (MpNCARF) and NON-CANONICAL IAA (MpNCIAA), are predicted to form part of the auxin-mediated transcriptional response (Flores-Sandoval et al., 2018; Mutte et al., 2018). NCARFs evolved from a class A-like ARF in the embryophyte ancestor, and in *M. polymorpha* act synergistically with MpARF1 to promote auxin signaling despite lacking a B3-DNA binding domain (Flores-Sandoval et al., 2018; Mutte et al., 2018). NCIAA genes form a sister clade to the AUX/IAA auxin signaling repressor genes, and as with the case of the class C ARFs, evolved prior to the origin of land plants (Flores-Sandoval et al., 2018; Mutte et al., 2018). Mpnciaa mutants are also sensitive to exogenous auxin (Mutte et al., 2018) suggesting they are not essential for auxin signaling. Despite conservation of these components across land plants, it is not known whether these genes act in common co-expression groups and whether their co-expression reflects functional dependency. Furthermore, the low number of annotated TFs, aMIR precursors and hormone-related genes in *M. polymorpha* allow accessible computation of a regulatory co-expression matrix, which perhaps is difficult in angiosperm model systems due to their high genetic redundancy. In this study we create a preliminary expression landscape of TFs defining developmental transitions in *M. polymorpha* based on available transcriptomic data. With co-expression analysis, we define gene clusters related to different tissues and



developmental stages. Moreover, we reveal a cluster of auxin-related genes that are validated by significant overlap with auxin-inducible transcripts. We further show that the class C auxin response factor is expressed independently of this cluster and may negatively influence reproductive transitions in *M. polymorpha*.

MATERIALS AND METHODS

DGE Expression Analysis

Fastq files of publicly available Illumina RNA-Seq libraries were obtained from the NCBI (SRA) depository. All libraries used for DGE analysis were performed in triplicates (except for antheridia – duplicates) and pairwise-comparisons were exclusively made with their original controls. *Marchantia* genome assembly (V3.1) and annotated gene models were used as references for TopHat2 mapping (Kim et al., 2013; Afgan et al., 2015). Control, *Mparf3* and *Mpmir160* transcriptomes were sequenced and processed as previously described (Flores-Sandoval et al., 2018). DGE analysis was performed with edgeR and outputs presented were limited to curated TFs (**Supplementary Table S1**; Bowman et al., 2017) with logCPM > 0 and *P*-Values < 0.01. Box plots were generated using an online R application available at <http://shiny.chemgrid.org/boxplotr/>. Accession numbers from the NCBI Sequence Read Archive include: 11-day thalli (DRR050343, DRR050344, and DRR050345), Archegoniophore (DRR050351, DRR050352, and DRR050353), Antheridiophore, (DRR050346, DRR050347, and DRR050348), Antheridia (DRR050349, and DRR050350), apical cell

(SRR1553294, SRR1553295, and SRR1553296), 13d-Sporophyte (SRR1553297, SRR1553298, and SRR1553299), Sporelings 0 h (SRR4450262, SRR4450261, and SRR4450260), 24hr-Sporeling (SRR4450266, SRR4450265, and SRR4450259), 48hr-Sporeling (SRR4450268, SRR4450264, and SRR4450263), 72hr-Sporeling (SRR4450267, SRR4450258, and SRR4450257), 96hr-Sporeling (SRR4450256, SRR4450255, and SRR4450254), 1-month wild type (SRR6685782, SRR6685783, and SRR6685784), 1-month *Mparf3* (SRR6685778, SRR6685779, and SRR6685785), 1-month *Mpmir160* (SRR6685777, SRR6685780, and SRR6685781), thallimock (SRR5905100, SRR5905099, and SRR5905098), and 2,4-D 1-h (SRR5905097, SRR5905092, and SRR5905091). Alternatively, we performed DGE analysis with DESeq2 (Love et al., 2014) when stated in the text.

Co-expression Analysis

Pearson's correlations were calculated for all *Marchantia* TFs and for genes with putative connection with auxin biology: correlation coefficients were calculated amongst TFs and whole genome co-expression partners were calculated only for auxin-related genes. Gene sets were generated considering those with correlation > 0.8 and *P*-value < 0.001 as significantly co-expressed. Venn and Euler plots were generated using R package VennDiagram as well as VENNY (Oliveros, 2007). UpSet plots were generated using its respective R package (Lex et al., 2014). Heatmaps for TF co-expression matrixes were generated using Heatmapper (Babicki et al., 2016) by applying average linkage clustering to rows and columns and Euclidian distance measurements.

Enrichment Analysis

Comparison of co-expression groups with differentially expressed genes were performed by enrichment analysis using Fisher's exact test. For the auxin enrichment analysis we manually selected auxin related co-expression groups (and unrelated genes as negative controls) as a bait against: Auxin inducible genes from *Physcomitrella patens* (p.adjust < 0.05; Lavy et al., 2016), *Arabidopsis thaliana* (p.adjust < 0.01; Omelyanchuk et al., 2017) and *M. polymorpha* (p.adjust < 0.05; Mutte et al., 2018). For the first two species listed we determined orthologs using Phytozome¹. $-\log_{10}$ P-values were reported in each Figure.

GO and Protein Family Enrichment

Gene sets were functionally characterized via GO term enrichment analysis using an in-house Fisher's exact test algorithm implemented in R (v3.3.2). We compiled Biological Process related GO term annotation for *M. polymorpha* genes using annotations available in Phytozome for *M. polymorpha* and *A. thaliana* and *P. patens* orthologs. We included in our GO term database both one-to-many and many-to-one relationships in order to obtain a more accurate annotation. For protein families, we used the same algorithm based on a different database. In this case, we annotated protein families using the HMMscan algorithm implemented in HMMER3 (Wheeler and Eddy, 2013) over the *M. polymorpha* proteome with default parameters and using Pfam profiles².

RESULTS

Transcription Factors Controlling Developmental Transitions in *M. polymorpha*

We performed DGE analysis on publicly available RNA-Seq libraries to (1) identify transcription factors defining or influencing *Marchantia* developmental transitions and (2) use the obtained datasets as references for enrichment analysis (Frank and Scanlon, 2015; Higo et al., 2016; Bowman et al., 2017). Datasets span multiple tissues throughout the life cycle, including intensively sampled sporelings during the first 5 days of development (Bowman et al., 2017), apical cells (Frank and Scanlon, 2015), gametangiophores, antheridia (Higo et al., 2016), and 13-day old sporophytic micro-dissections (Frank and Scanlon, 2015). In our analysis, 82% of differentially expressed TFs with $P < 0.01$ (edgeR) have fold changes (FC) below 4× in most tissues except sporophytes (Figure 2A and Supplementary Figure S1A), suggesting feedback regulation preventing drastic changes in regulatory gene expression. To account for both read abundance and fold-changes, we used the product of logFC and logCPM (counts per million) to categorize the transcripts enriched in a particular tissue. Of the 405 TFs in the *M. polymorpha* genome (Supplementary Table S1), 45 were

not differentially expressed in any of the performed pair-wise comparisons (Supplementary Figure S1B).

In order to find TFs specific to a developmental transition, we compared all upregulated TFs in each tissue using a cut off of logFC and logCPM > 0 (Figure 2B and Supplementary Table S2). Mature sporophyte-enriched TFs sequenced by JGI (Bowman et al., 2017) were included in this comparison to validate the 13-day microdissected sporophytes with both datasets showing a high degree of overlap (Figure 2B). Mature sporophytes have the largest number of uniquely upregulated TFs (Figure 2B), followed by 24-h germinated spores, gametangiophores and thalli. The apical cell and antheridia had a single uniquely upregulated TF each (Figure 2B). As expected, antheridia and antheridiophores, male and female gametangiophores and thalli/24-h sporelings show the highest degree of similarity. Surprisingly, 13-day sporophytes, but not mature sporophytes, had common and exclusive upregulation with archegoniophores and antheridia. Finally, antheridia, antheridiophores and archegoniophores formed a large group of uniquely upregulated TFs (Figure 2B). The TFs enriched in each tissue library will be described in detail in the following paragraphs.

Sporelings 0 to 24 h

A total of 193 TFs (114 upregulated) define the transition of dormant spores to germinated ones (Figure 2A and Supplementary Table S3). This stage has the third highest number of upregulated TFs after gametangiophores (logFC > 0), and has the highest number of upregulated TFs at logFC > 2 (Supplementary Figure S1A). This likely reflects the large number of physiological, cellular and developmental processes being activated upon germination. The ortholog of the root specifying *TARGET OF MONOPTEROS 5/ABNORMAL SHOOT 5* (Schlereth et al., 2010), MpbHLH7 (Mapoly0039s0068, logFC = 8 and logCPM = 5.9) is the highest expressed TF incorporating both fold-changes and read abundance in 1-day sporelings (Figure 3A). MpbHLH7 is followed by the B-box type ZINC FINGER gene MpBBX3 (Mapoly0049s0067, logFC = 8.4 and logCPM = 5.7) and the single *M. polymorpha* GROWTH REGULATING FACTOR gene MpGRF (Mapoly1350s0001, logFC = 10 and logCPM = 4.4). MpGRAS8 (Mapoly0065s0089, logFC = 7.6 and logCPM = 5.23) from an uncharacterized GRAS lineage in plants (Bowman et al., 2017), and MpbHLH41 (Mapoly0100s0033, logFC = 6.21 and logCPM = 6.23), an ortholog of NAI1, a TF involved in activation of post-germinative seedling programs in *Arabidopsis* (Yoshii et al., 2015) are also highly upregulated. An auxin-inducible gene (Mutte et al., 2018), the HOMEODOMAIN class II HD-ZIP, MpC2HDZ (Mapoly0069s0069, logFC = 5.89 and logCPM = 5.01) also has high expression consistent with up-regulation of MpYUC2 (Mapoly0063s0040, logFC = 1.32 and logCPM = 3.55). Co-expression dynamics for day-1 sporeling enriched genes shows MpbHLH7 and MpC2HDZ are highly correlated. Meanwhile MpGRF expression correlates with that of the class I TCP MpTCP1, the auxin signaling repressor MpIAA and the HOMEODOMAIN class IV HD-ZIP MpC4HDZ (Supplementary Figure S2).

¹<https://phytozome.jgi.doe.gov/>

²<https://pfam.xfam.org/>

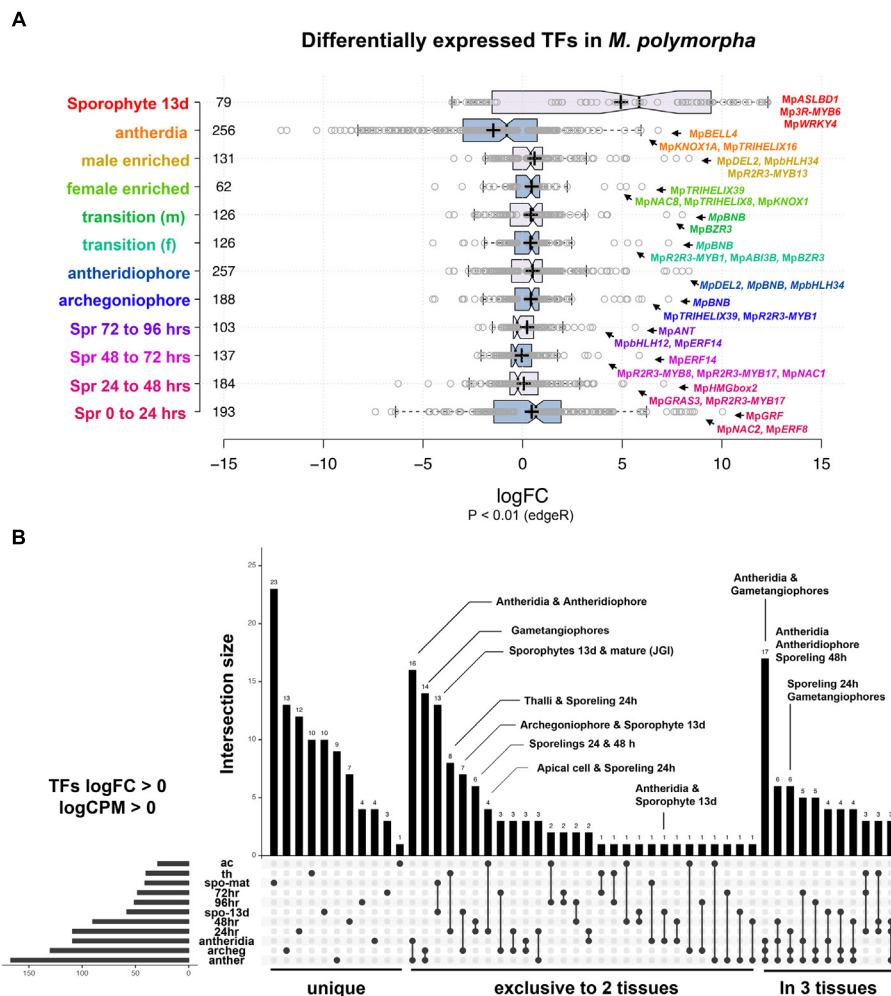


FIGURE 2 | Differential gene expression analysis in *M. polymorpha*. **(A)** Boxplots of \log_2FC value distribution of differentially expressed (edgeR, $P < 0.01$) TFs in multiple pair-wise comparisons. Central lines represent medians; box limits indicate the 25th and 75th percentiles as determined by R software (<http://shiny.chemgrid.org/boxplot/>); whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles, outliers are represented by dots; crosses represent sample means; data points are plotted as open circles. Number of sample points are included for each group. **(B)** UpSet diagram indicating intersections between upregulated TFs ($P < 0.01$, $\log_2FC > 0$) in multiple pairwise comparisons. Diagram indicates uniquely up-regulated TFs on the left, followed by those exclusively shared between two and three pairwise comparisons. Abbreviations: ac, apical cell; th, thallus; spo-mat, JGI-sporophytes; 72 h, 72-h sporelings; 96 h, 96-h sporelings; spo-13 d, 13-day sporophytes; 48 h, 48-h sporelings; 24 h, 24-h sporelings; archeg, archegoniophores; anther, antheridiophores.

Sporelings 24 to 48 h

A total of 184 TFs (90 upregulated, **Supplementary Table S4**) define the second day of sporeling development where the first rhizoid is formed (Bowman et al., 2017). *MpHMGbox2* (Mapoly0026s0138, $\log_2FC = 7$ and $\log_2CPM = 6.8$) and *MpHMGbox1* (Mapoly0013s0048, $\log_2FC = 1.03$ and $\log_2CPM = 8.71$) are upregulated and abundant (**Figure 3B**). They are orthologs of the 3xHIGH MOBILITY GROUP-BOX2 genes involved in cell cycle regulation (Pedersen et al., 2011). These are followed by two GRAS TFs: *MpGRAS3* (Mapoly0014s0183, $\log_2FC = 5$ and $\log_2CPM = 3.2$) and *MpGRAS5* (Mapoly0047s0034, $\log_2FC = 2.86$ and $\log_2CPM = 3.85$), which belong to separate and uncharacterized GRAS lineages in land plants (Bowman et al., 2017). *MpERF21* (Mapoly0166s0010, $\log_2FC = 2.77$ and $\log_2CPM = 4.38$), an ABA REPRESSOR1

(ABR1) ortholog and *MpERF7* (Mapoly0034s0029, $\log_2FC = 1.77$ and $\log_2CPM = 4.9$) a DREB26 ortholog, are two AP2/ERF TFs enriched in this transition and putatively involved in stress-responses (Pandey, 2005; Krishnaswamy et al., 2011). Finally, *MpbHLLH15* (Mapoly0038s0058, $\log_2FC = 2$ and $\log_2CPM = 6$), an ortholog of CRYPTOCHROME-INTERACTING (CIB) bHLHs (Liu et al., 2008) is also enriched in day two sporelings (**Figure 3B**). Co-expression dynamics for day-2 sporeling enriched genes shows *MpHMGbox1* and 2 form a distinct group, while *MpERF7* and 21 are co-expressed (**Supplementary Figure S3**).

Sporelings 48 to 72 h

A total of 142 TFs (50 upregulated, **Supplementary Table S5**) define this transition, where photosynthetic vs. anchoring

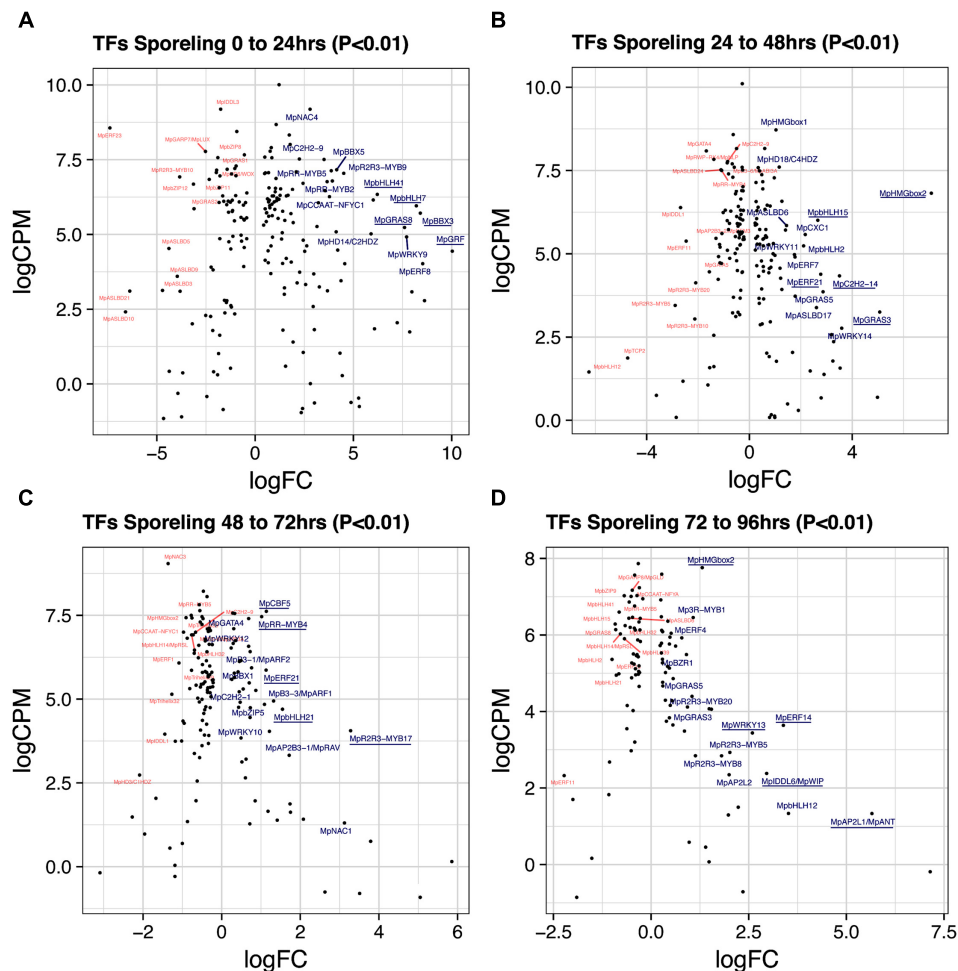


FIGURE 3 | Differentially expressed TFs in *M. polymorpha* sporlings. LogFC and logCPM outputs of DGE analysis showing pairwise comparisons of (A) 0 vs. 24 h, (B) 24 vs. 48 h, (C) 48 vs. 72 h, and (D) 72 vs. 96 h. Top 15 upregulated (blue) and downregulated (red) TFs are annotated, with the top5 upregulated TFs underlined.

tissue polarity is properly established (Bowman et al., 2017). *MpR2R3-MYB17* (logFC = 3.2, logCPM = 4.05) a MIXTA-like MYB TF is the highest expressed gene after incorporating abundance (Figure 3C). It belongs to a MYB lineage (clade IV) composed of genes regulating secondary compound metabolism (Bowman et al., 2017). Similar to day 2, *MpERF21* continues its enrichment (logFC = 1.12 and logCPM = 5.86), as well as *MpbHLH21* (Mapoly0037s0007, logFC = 1.54 and logCPM = 4.69) a paralog of *MpbHLH15* (CIB bHLHs). The auxin signaling pathway, represented by *MpARF1* (Mapoly0019s0045, logFC = 1.32 and logFC = 4.94) and *MpARF2* (Mapoly0011s0167, logFC = 0.72 and logFC = 6.41) are significantly upregulated in this transition (Supplementary Figure S4A). The CUP-SHAPE COTYLEDON ortholog *MpNAC1* (Mapoly0015s0058, logFC = 3.1 and logCPM = 1.3) and the RAV-like B3-domain TF *MpRAV* (Mapoly0072s0102, logFC = 1.71 and logCPM = 3.32) are also enriched. The logFC average in 3-day sporlings has decreased by $\sim 1/2$ (Figure 2A) from 1-day sporlings, suggesting increased homeostatic control of TF expression. *MpR2R3-MYB17*, *MpARF1*, *MpARF2*, and

MpNAC1 form a robust co-expression cluster (Supplementary Figure S5).

Sporlings 72 to 96 h

A total of 106 TFs (52 unregulated, Supplementary Table S6) define this transition, where there is a proliferation of photosynthetic cells and elongation of the rhizoid (Bowman et al., 2017). The *ANT/PLETHORA/BABYBOOM* ortholog *MpANT* (Mapoly0008s0071, euANT clade) involved in meristematic activity in mosses (Aoyama et al., 2012) and angiosperms (Aida et al., 2004; Kareem et al., 2015) is the TF with the highest fold-change (logFC = 5.6) despite being mildly expressed (logCPM = 1.3, Figure 3D). The auxin-induced gene *MpWIP* (Mapoly0096s0050, logFC = 2.95 and logCPM = 2.38) necessary for air pore development (Jones and Dolan, 2017) follows a similar pattern. *MpERF14* (Mapoly0066s0103, logFC = 3.38 and logCPM = 3.63) from clade VIII of the ERF phylogeny (Bowman et al., 2017), *MpHMGbox2* (logFC = 7.75 and logCPM =) and *MpWRKY13* (Mapoly0162s0003, logFC = 2.59 and logCPM = 3.43) are highly expressed and upregulated

(Figure 3D). The auxin-signaling TFs MpARF1 (logFC = 0.46 and logCPM = 5.76) and MpARF2 (logFC = 0.24 and logCPM = 6.91) continue to be highly expressed and upregulated (Supplementary Figure S4A). Enrichment analysis (Fishers-exact test) probing auxin-induced *M. polymorpha* transcripts (Mutte et al., 2018) with sporeling and thalloid transcriptomes shows that 96-h sporelings vastly converge with genes activated by auxin, perhaps as a result of progressive increases of class A and B ARF activity (Supplementary Figure S4B). All of the above mentioned genes form a robust co-expression cluster (Supplementary Figure S5).

Thalli vs. Archegoniophores

A total of 188 TFs (129 upregulated, Supplementary Table S7) are differentially expressed between 11-day thalli and mature archegoniophores (Higo et al., 2016). The TFs with highest expression at this stage (Figure 4A) is the bHLH factor MpBONOB/MpBNB (Mapoly0024s0106, logFC = 7 and logCPM 2.6), which has been shown to promote formation of gametangiophores and biogenesis of gametes (Yamaoka et al., 2018). MpR2R3-MYB1 (Mapoly0001s0061) is a highly expressed and upregulated factor in archegoniophores (logFC = 5.26 and logCPM = 6.37), followed by the B3-domain MpABI3B (Mapoly0474s0001, logFC = 5.26 and logCPM = 5.96) whose orthologs in *Arabidopsis* modulate abscisic acid responses (McCarty et al., 1991; Suzuki et al., 2001). Finally, the putative GT2-like MpTRIHILIX39 (Mapoly0003s0229, logFC = 5.98 and logCPM 3.2) has drastic fold changes but lower read counts (Figure 4A).

Thalli vs. Antheridiophores

A total of 257 TFs (166 upregulated, Supplementary Table S8) are differentially expressed in antheridiophores when compared to 11-day thalli. One of the two DP-E2F-LIKE PROTEIN TFs in *Marchantia*, MpDEL2 (Mapoly0198s0008, logFC = 8.34 and logCPM = 4.64), involved in regulating the cell cycle and cell proliferation (Sozzani et al., 2010), as well as MpBNB (logFC = 8 and logCPM = 3.4) are the TFs with the highest fold change in the transition from thalli to antheridiophores (Figure 4B). In addition, MpERF21 (Mapoly0166s0010, logFC = 5.14 and logCPM = 6.52), MpBZR3 (Mapoly0072s0031, logFC = 7.22, logCPM = 4.23), and MpTRIHILIX34 (Mapoly0159s0017, logFC = 4.38, logCPM = 6.84) are highly expressed genes with large fold-changes in antheridiophores (Figure 4B).

Reproductive Transition

A total of 126 TFs (87 upregulated, Supplementary Table S9) are shared with similar dynamics in thalli vs. archegoniophore and thalli vs. antheridiophore comparisons (Figure 4C). Nine of these genes were enriched in no other tissue but gametangiophores albeit being mildly upregulated (Figure 2B and Supplementary Table S2). Thirteen genes were exclusive to gametangiophores and antheridia, corroborating their specificity in reproductive roles (Figure 2B and Supplementary Table S2). MpBNB is in this former group, consistent with its role in both gametangiophore and gamete formation (Yamaoka et al., 2018). Additional genes were enriched but not specific to

gametangiophores (Supplementary Table S2). A majority of these TFs (Figure 4C) have 1:1 ratios of logCPM between male and female tissues with only MpR2R3-MYB1 being higher in females (logCPM female/male = 1.3) and MpBZR3 and MpR2R3-MYB6 preferentially expressed in males (logCPM male/female = 2.5 in both cases). Despite their high expression in males, complete MpR2R3-MYB1 transcripts were exclusively found in females, suggesting male expression may be represented by incomplete transcripts.

Female Enriched Genes

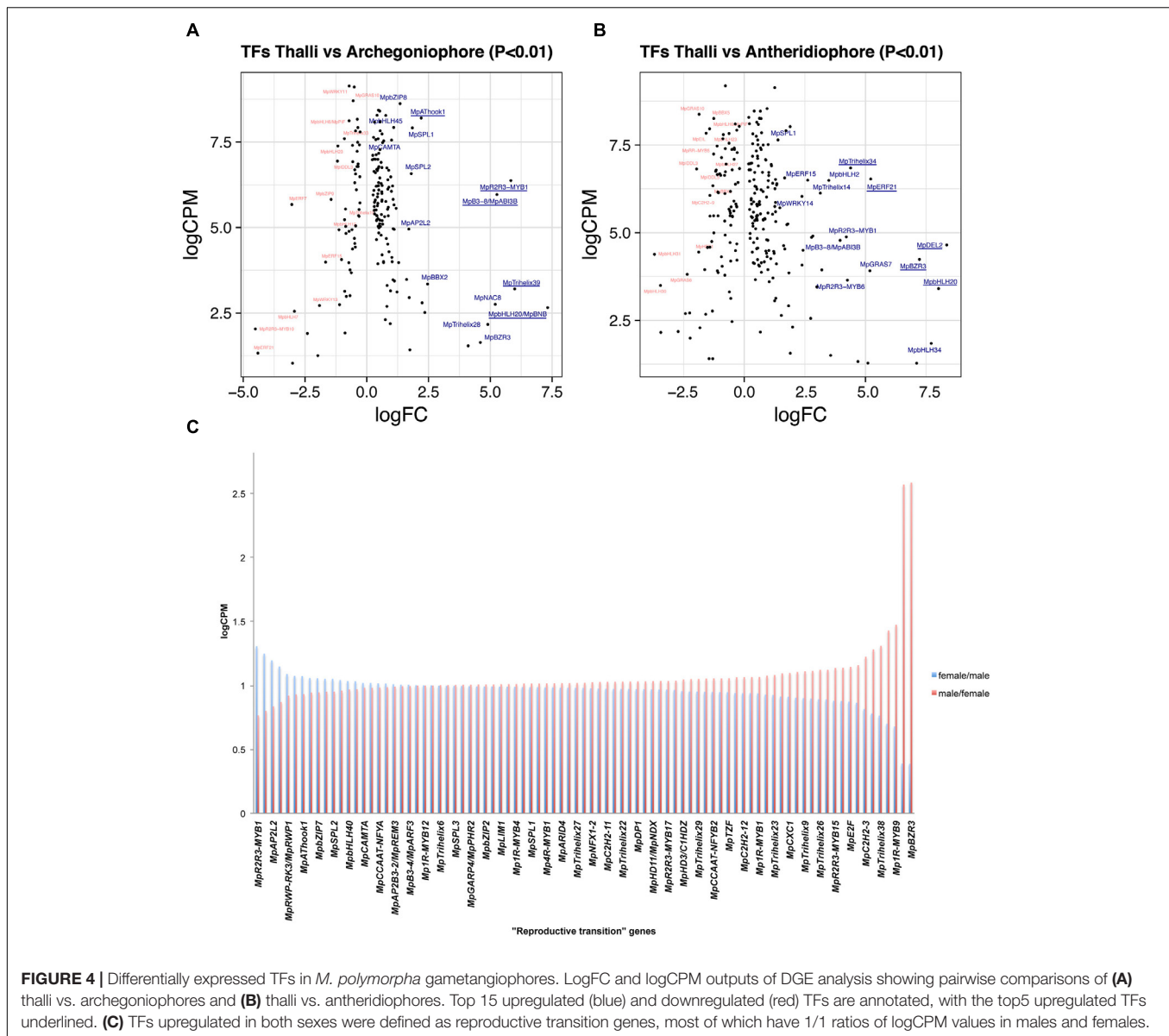
Due to the lack of archegonia-specific transcriptomes, we defined female enriched genes by discarding reproductive transition genes from the thalli vs. archegoniophore comparison (62 genes in total, 42 upregulated, Supplementary Table S10). These include (Figure 5A) MpNAC8 (Mapoly0035s0054, logFC = 5.19 and logCPM = 2.75), which is an ortholog of LONG VEGETATIVE PHASE (LOV1) involved in repressing flowering transitions in *Arabidopsis* (Yoo et al., 2007); MpTRIHILIX39, and the SH4-like MpTRIHILIX28 (Mapoly0122s0030, logFC = 4.89 and logCPM = 2.16). Additionally, two TALE-HOMEODOMAIN TFs are also upregulated in females although with lower read counts, possibly showing a diluted signal from archegonia RNA. These include the class I KNOX gene MpKNOX1 (Mapoly0175s0020, logFC = 4.1 and logCPM = 1.53) as well as a BELL gene, MpBELL5 (Mapoly0093s0028, logFC = 2.23 and logCPM = 2.79), the latter only having orthologs in charophycean algae (Bowman et al., 2017).

Male Enriched Genes

A total of 131 TFs (79 upregulated, Supplementary Table S11) are enriched in antheridiophores after discarding reproductive transition genes. MpDEL2 and MpTRIHILIX34 are kept in this category (Figure 5B). Additional genes include, MpERF21 (Mapoly0166s0010, logFC = 5.21 and logCPM = 6.52), MpHLLH2 (Mapoly0028s0062, logFC = 3.48 and logCPM = 6.49) an ortholog of AT4G09820, MpERF15 (Mapoly0068s0088, logFC = 2.61 and logCPM = 6.49), MpGRAS7 (Mapoly0064s0023, logFC = 5.18 and logCPM = 3.91) a gene lost in the common ancestor of seed plants (Bowman et al., 2017), and MpWRKY14 (Mapoly0467s0001, logFC = 2.37 and logCPM = 6.49) that belongs to clade III in the WRKY phylogeny (Bowman et al., 2017).

Antheridiophore vs. Antheridia

A total of 256 TFs (108 upregulated, Supplementary Table S12) are differentially expressed in antheridia when compared to antheridiophores. The higher ratio of repressed/activated TFs suggests the substitution of haploid photosynthetic and free-living thalli for that of more specialized gamete forming pathways (Figure 2A). For example, TFs upregulated during sporeling development, such as MpGRF (logFC = -10) and MpANT (logFC = -8) are repressed (Figure 5C). Upregulated genes in antheridia in both in terms of fold-change and abundance (Figure 5C) are an ortholog of the SH4-like MpTRIHILIX16 (Mapoly0037s0146, logFC = 5.79 and



logCPM = 7.99); MpKNOX1A (Mapoly0174s0007, logFC = 5.93 and logCPM = 7); and MpKNOX1B (Mapoly0023s0081, logFC = 2.55 and logCPM = 4.37), which are semi conserved TALE (KNOX1) HD genes that have lost the DNA binding domain (Bowman et al., 2017), MpBELL4 (Mapoly0013s0027, logFC = 6.79 and logCPM = 4.75) and MpBELL5 (logFC = 2.25 and logCPM = 2.27). MpBELL4 in particular is not detected as differentially expressed in thalli vs. antheridiophores, suggesting it is male-gamete specific, while MpBELL5 shows incomplete transcription in males, suggesting that it might be a female-associated factor. Additional genes previously curated (Higo et al., 2016) include Mp1R-MYB22 (Mapoly0811s0001, logFC = 3.85 and logCPM = 7.15), MpRWP2/MpMID (Mapoly0014s0044, logFC = 3.83, logCPM = 7) whose ortholog in the chlorophycean algae *Chlamydomonas* specifies minus gametes

(Ferris and Goodenough, 1997; Lin and Goodenough, 2007), and MpR2R3-MYB18 (Mapoly0123s0012, logFC = 3.52 and logCPM = 6.81) an ortholog of FOUR LIPS (FLP), which restricts mitosis in stomatal development in *Arabidopsis* (Lee et al., 2013).

Sporophytic Genes

We re-analyzed the microdissected RNA of 13-day-old sporophytes given that at the time of publication (Frank and Scanlon, 2015) there was no annotated reference genome. These samples have low read-mapping rate using TopHat2 (~60%), which would leave differentially expressed genes with low CPM undetected. Replicates were compared to its endogenous control, the apical notch microdissected libraries. Due to these limitations, the 85 differentially expressed TFs (57 upregulated, **Supplementary Table S13**) in sporophytes are skewed to high-fold changes (**Figures 2A, 5D**).

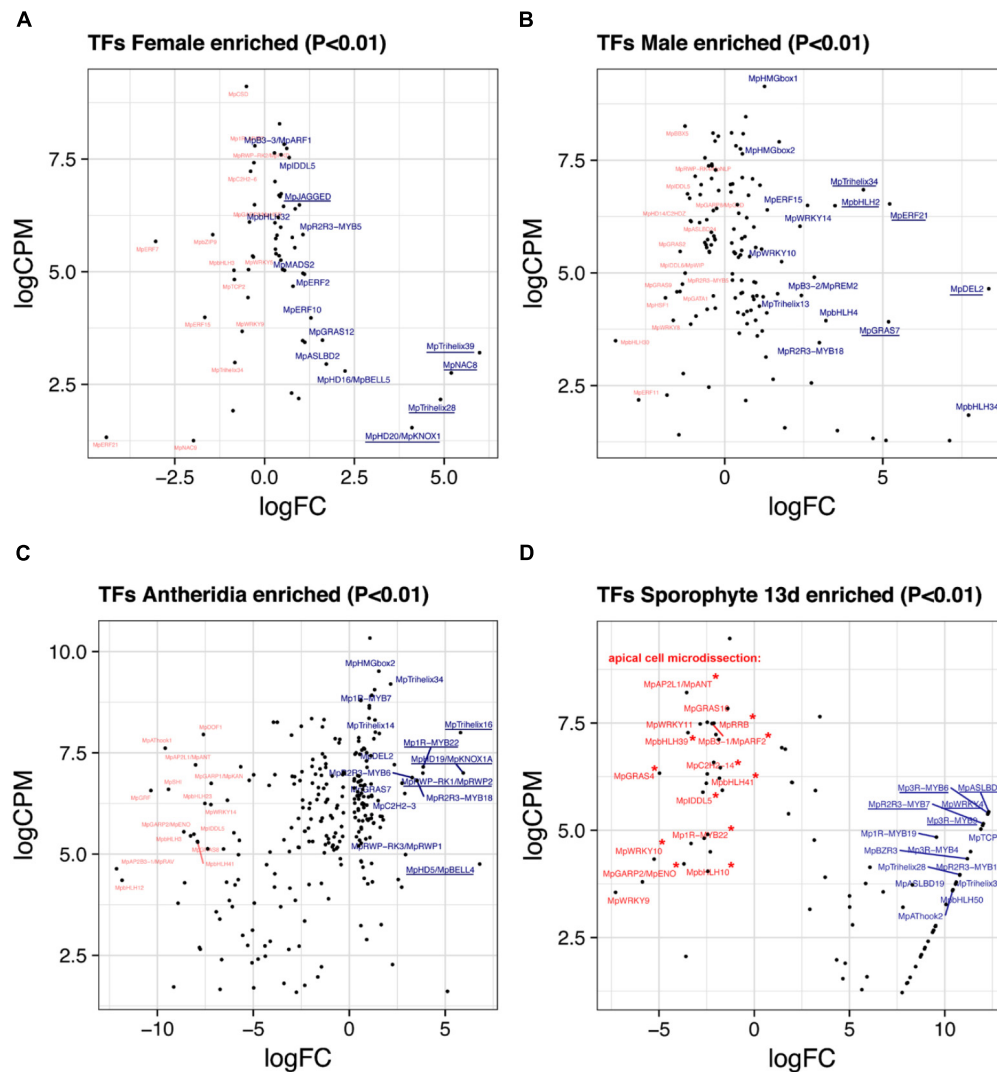


FIGURE 5 | Sex-enriched and sporophyte TFs obtained from DGE analysis. LogFC and logCPM outputs of DGE analysis showing TFs solely enriched in (A) archegoniophores and not antheridiophores, (B) antheridiophores and not archegoniophores, (C) antheridia vs. antheridiophores, and (D) 13-day sporophytes vs. apical cell RNAs. Top 15 upregulated (blue) and downregulated (red) TFs are annotated, with the top 5 upregulated TFs underlined. Asterisks in (D) represent TFs not upregulated in thallus tissue.

The TFs shared between 13-day sporophytes and mature JGI sporophyte libraries (Figure 2B) include the AP2/B3 domain MpREM4 (Mapoly0006s0205, logFC = 13.86 and logCPM = 6.96), several FAMA-like orthologs MpbHLH35-37 and 50 (Mapoly0031s0072, 74, 75, and 76; average logFC = 9.4 and logCPM = 2.67) involved in stomata development in mosses and tracheophytes (Ohashi-Ito and Bergmann, 2006; Chater et al., 2016); multiple Mp3R-MYBs (Mp3R-MYB3-6 and 9) among others (Supplementary Table S13). However, representative mature sporophyte-enriched genes such as the TALE-class HOMEODOMAIN MpBELL1 (Mapoly0213s0014), MpKNOX2 (Mapoly0194s0001) and the AP2/B3 domain MpREM1 (Mapoly0183s0010) were not upregulated in 13-day sporophytes (Supplementary Table S13). Enrichment analysis confirmed significant overlap between differentially expressed

JGI and 13-day sporophyte transcriptomes with a slight significant overlap between female-enriched genes and 13-day sporophytes (Supplementary Figure S6A). Female enriched TFs (Figure 5A) upregulated in 13-day sporophytes include MpTRIHELIX28 and 39, MpASLBD2, MpKNOX1, MpGRAS12, MpNAC8, and MpbHLH29 (Figure 2B and Supplementary Table S2). This could be due to archegonia RNA contamination or a continuation of maternal developmental programs in young sporophytes. Supporting the latter interpretation, 13 of previously classified antheridia enriched genes (Figure 5C) including MpBELL4, MpTRIHELIX16, MpR2R3-MYB3, 6 and 13, MpGRAS7 and MpDEL2 were also detected as upregulated in 13-day sporophytes (Supplementary Figure S6B). MpBELL4 is the only TF exclusively upregulated in antheridia and 13-day sporophytes in our analysis (Supplementary Table S2).

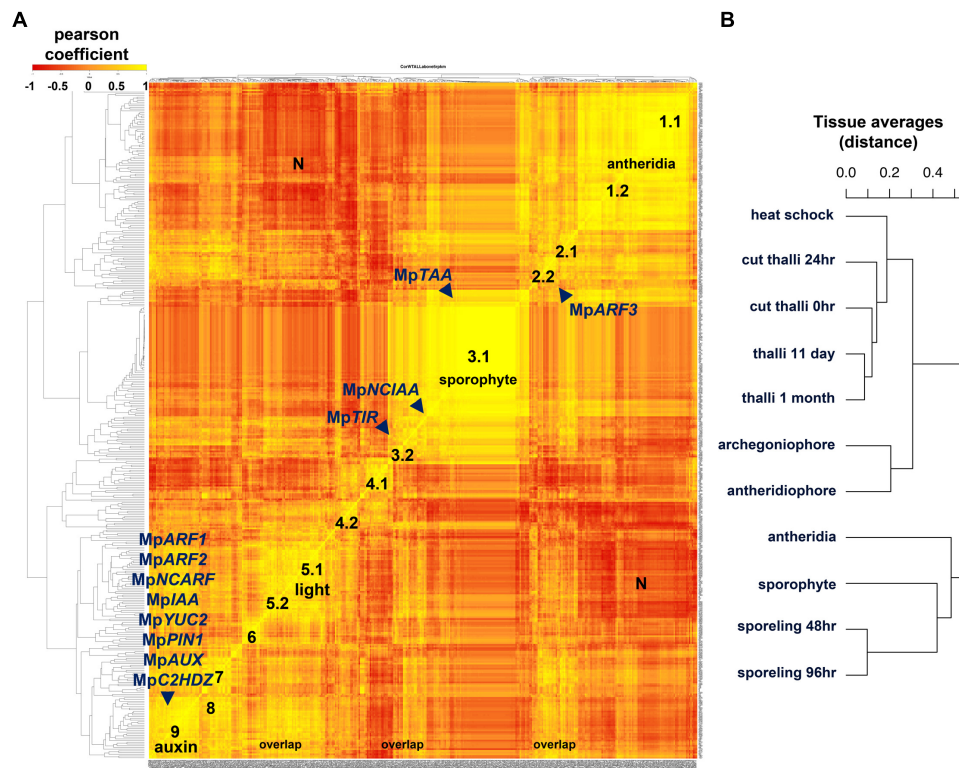


FIGURE 6 | Co-expression matrix of regulatory genes in *M. polymorpha* (A) Heatmap of Pearson coefficient matrix of all annotated *M. polymorpha* TFs, MIR precursors and putative hormone response and synthesis pathway genes (Supplementary Table S15) in 11 tissue libraries using RPKM values > 5. Co-expression groups were delimited by sets of genes sharing coefficients above 0.8. Genes putatively involved in auxin biosynthesis, perception, signaling and transport are mapped. (B) Distance dendrogram of tissue libraries used in the analysis. Average replicate RPKM values were used per library. Scale for heatmap indicates Pearson coefficients.

Finally, MpBELL5 is upregulated in archegoniophores, antheridia (incomplete transcript) and 13-day sporophytes (Supplementary Table S2).

There are additionally seven genes uniquely upregulated in 13-day sporophytes (Figure 2B), these include: MpWRKY4 (Mapoly0031s0170), MpASLBD19 (Mapoly0817s0001), MpAP2L5 (Mapoly0144s0032), among others (Supplementary Table S2). Orthologs of sporophyte-enriched *Physcomitrella* genes (Ortiz-Ramírez et al., 2016) include (Supplementary Table S2) the class II TCP TF MpTCP2 (Mapoly0001s0298, logFC = 11.89 and logCPM = 5.02), MpREM4, MpABI3B (Mapoly0474s0001, logFC = 8.29 and logCPM = 3.72) and the SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE MpSPL1 (Mapoly0014s0224, logFC = 1.44 and logCPM = 6.92). These genes are upregulated in other tissues and only MpREM4 seems to be sporophyte-specific (Supplementary Table S2).

Apical Cell

We used the microdissected putative apical cell libraries as controls for the sporophyte DGE analysis (Frank and Scanlon, 2015). This potentially shows that down-regulated genes in sporophytes are upregulated in apical cells (Figure 5D). To try to refine this comparison we discarded genes that were robustly (logFC > 1) upregulated in other tissues

(Supplementary Figure S7 and Supplementary Table S14). This resulted in 8 TFs exclusively enriched at logFC > 1 in apical cells. These included a GARP TF MpGARP2/ENO (Mapoly0157s0009, logFC = 5.87 and logCPM = 3.79) belonging to a lineage sister to KANADI genes (Bowman et al., 2017), the B3-domain MpARF2 auxin signaling transcriptional repressor (Mapoly0011s0167, logFC = 2.2 and logCPM = 7.48), the single member of the basal SCGJ bZIP group MpbZIP15 (Mapoly0737s0001, logFC = 2.13 and logCPM = 6.58) and 5 other genes (Supplementary Table S4). Outside this group, the highest expressed TF in the apical cell RNA is MpGRAS4 (Mapoly0031s0041, logFC = 4.97 and logCPM = 6.32), an ortholog of LATERAL SUPPRESSOR involved in axillary meristem establishment in *Arabidopsis* (Greb et al., 2003). MpGRAS4 however, is strongly upregulated in 24 h-sporelings (logFC = 3.26 and logCPM = 3.24), 48 h-sporelings (logFC = 1.18) and weakly in antheridiophores (logFC = 0.89). MpANT is robustly enriched in apical cell tissue (logFC = 3.55 and logCPM = 8.21) but also in 96 h sporelings (Figure 2A). Four genes were exclusively upregulated in apical cells and 24 h-sporelings at logFC > 1: MpbHLH39 (Mapoly0073s0059, logFC = 3.48 and logCPM = 7.27), MpbHLH41, the type-B RESPONSE REGULATOR MpRR-B (Mapoly0101s0006) and MpWRKY9 (Mapoly0051s0057). Finally, MpWRKY10 (Mapoly0057s0012, logFC = 5.25 and logCPM = 4.32) is

upregulated exclusively in apical cells and antheridiophores (**Supplementary Table S2**).

Marchantia Regulatory Genes Are Organized in Co-expression Groups

In order to predict functional roles of clusters of TFs sharing similar expression dynamics, we created a co-expression matrix comparing averages of normalized reads per kilobase per million (RPKM) across 11 different tissues. Pearson coefficients were calculated for each of the annotated *Marchantia* TFs, microRNA (miR) precursors as well as hormone biosynthetic and signaling genes. Two stages of sporeling development (48 and 96 h), multiple thallus RNAs, archegoniophores, antheridiophores, antheridia and older sporophytes (JGI) were used in this analysis. Using a minimal RPKM value of 5 as a threshold, a Pearson coefficient heatmap (**Figure 6A**) suggests at least 9 highly correlated groups, some of which show a greater degree of correlation than others. A distance dendrogram (RPKM < 1) supported the occurrence of multiple groups (**Supplementary Figure S8**) and we focused on describing nine of the most conspicuous in basis of their tissue-specificity or putative functional relationships (**Supplementary Figure S9**). The antheridia (group 1) and sporophyte-enriched genes (group 3) combined comprise ~1/2 of the total sample. This is in agreement with these tissues being the most divergent in terms of form and physiology.

Group 1 defined by MpBELL4 partners (Pearson > 0.8, average coefficient = 0.91) composed of 47 TFs includes MpTRIHILIX16, MpKNOX1A, MpKNOX1B, MpRWP2/MID, Mp1R-MYB22, MpPIN2, MpLFY, MpGRAS7 among others (**Supplementary Figure S9A**). A diffuse group (group 2) includes 9 partners of MpARF3 (Pearson > 0.8, average coefficient = 0.85), many of which were classified as reproductive transition genes, such as, MpSPL1-3, MpbZIP8, MpbHLH17 and 48, MpRR-B, MpC3HDZ, among others (**Supplementary Figure S9B**). Sporophytic genes (group 3) can be divided between those enriched or specific to the sporophyte. The MpBELL1 subgroup (Pearson > 0.8 and average coefficient = 0.94) is composed of 96 genes including MpREM1 and 4, MpKNOX2, MpYUC1, MpPYL2-4, Mp1R-MYB5,6,19 and 21; Mp3R-MYB2,4-9; MpbHLH8, 9, 11, 18, 22, 25, 34-37 and 51; MpASLBD1,-5, 8, 9, 11, 13-16, 19 and 21, most of which are exclusive to the mature sporophyte (**Supplementary Figure S9C**). Meanwhile, the MpNCIAA subgroup (Pearson > 0.8, average coefficient = 0.84) is composed of 25 genes including MpTAA, MpERF17, MpGRAS2, MpLUX, MpPHR2, MpACS2, MpMIR319a, MpIDDL1-2, MpASLBD2 and others that fit in the first category (**Supplementary Figure S9D**). A diffuse group 4.1 (Pearson > 0.8, average coefficient = 0.88) includes six partners of MpbHLH14/MpRSL (**Supplementary Figure S9E**), which has been shown to promote epidermal outgrowths (Proust et al., 2016). Members of this group include MpNAC4, MpGRAS8, and MpMIR11671 a negative regulator of MpSPL1 (Tsuzuki et al., 2015; Lin et al., 2016). Group 4.1 is expressed maximally in 48 h. sporelings and steadily decreases at later stages (**Supplementary Figure S9E**). Similar to group 4.1, MpGLD forms a subgroup

4.2 (Pearson > 0.8, average coefficient = 0.91) with five other genes including MpC2H2-9, MpbHLH32, 41, and MpRR-MYB5 that follows a similar decrease in expression pattern albeit with higher expression in the thallus (**Supplementary Figure S9F**). Group 5 is defined by MpTOC1 partners (Pearson > 0.8, average coefficient = 0.81), composed of 12 TFs including MpbHLH6/MpPIF, the Type-A RESPONSE REGULATOR MpRR-A, MpBBX5, MpWIP, MpETR1, MpERF10, MpGRAS9 among others, possibly reflecting a light signaling/circadian clock module (Inoue et al., 2016; Linde et al., 2017). Group 5 has peaks in thalli with excised apical notches (0 h) and heat-shocked thalli (**Supplementary Figure S9G**). MpKAN and its partners overlap with most of group 5 (Pearson > 0.8, average coefficient = 0.86) but form a larger group with 42 other members that include MpWRKY8 and 11, MpbHLH15, 23, 27 and 31, MpHSR, and auxin-related genes (which will be discussed in the following paragraph) at the ~20th rank. The MpKANADI group has its average highest expression only at 0-h cut thalli with excised apical notches suggesting a role in differentiated tissues (**Supplementary Figure S9H**). Group 6 (**Supplementary Figure S9I**) comprises 10 genes highly expressed in regenerating thalli (after 24 h of apical notch excision) and includes partners of the apical-cell enriched MpGRAS4 (Pearson > 0.8, average coefficient = 0.85). Other partners include MpR2R3-MYB5 and 20, MpERF20, MpbHLH4, MpGRAS6/DELLA and MpLOG. Given their expression context it is likely that this group regulates certain aspects of thaloid totipotency/regeneration. Group 7 genes are MpSPL2 partners (Pearson > 0.8, coefficient average = 0.88) that overlap considerably with the MpARF3 group, it also includes novel genes such as MpATHook1, MpPIN3, MpMIR11697 which targets MpYUC2 (Lin et al., 2016), MpPLINC and MpGRAS12 (**Supplementary Figure S9J**). Group 8 (Pearson > 0.8 and coefficient average = 0.87) is defined by partners of MpNAC8 and includes 16 putative female enriched TFs (**Supplementary Figure S9K**) such as MpKNOX1, MpBPC1, MpTRIHILIX28 and 39, MpPIN3, and MpGRAS12. Group 8 has overlaps with group 2/7 members such as MpSPL2, MpABI3B, MpMIR11697 and it includes MpR2R3-MYB1 a reproductive transition gene with highest expression in females (**Supplementary Figure S9K**).

Surprisingly, most auxin-related genes formed a large co-expression group (group 9) of 32 TFs (Pearson > 0.8, average coefficient = 0.87, **Figure 6**). Group 9 is defined by partners of the class A ARE, MpARF1 (**Supplementary Figure S9L**), a TF regulating physiological and transcriptional auxin responses, gemmae development and gemmae dormancy (Kato et al., 2017). Additional members include the class B ARE, MpARF2 (**Supplementary Figure S10**), which has been identified as a transcriptional repressor (Kato et al., 2015), the auxin signaling repressor MpIAA (Flores-Sandoval et al., 2015; Kato et al., 2015), the NON-CANONICAL ARF MpNCARF (Mutte et al., 2018), MpYUC2 an auxin biosynthetic gene (Eklund et al., 2015), MpPIN1 a putative intercellular auxin exporter (Bowman et al., 2017), MpAUX/LAX an intercellular auxin importer (Bowman et al., 2017), MpC2HDZ an auxin responsive gene (Mutte et al., 2018), and MpSHI a putative

YUC2 activator (Eklund et al., 2010), as well as additional TFs involved in sporeling patterning such as MpGRF and MpANT. A cluster of bHLH genes including MpbHLH10, 42, 43, 45, 47, and 48 are all members of this group. In agreement with recent works (Flores-Sandoval et al., 2018; Mutte et al., 2018) MpARF3 is not co-expressed with any member of the auxin group except for MpNCARF (Pearson = 0.78). A third set of putative auxin related genes showing independent co-expression dynamics includes the putative auxin receptor MpTIR1 as well as NON-CANONICAL IAA MpNCIAA, which are associated with sporophytic genes (**Supplementary Figure S9D**). The dn-OPDA receptor in *Marchantia* MpCOI (Monte et al., 2018), is co-expressed with MpARF1 (Pearson = 0.88). MpTAA the major catalyst of Tryptophan into IPA, a precursor of auxin in *M. polymorpha* (Eklund et al., 2015) is also preferentially expressed in older sporophytes grouping with the sporophyte-specific MpYUC1 (Pearson = 0.91, **Figure 6** and **Supplementary Figure 9D**).

Testing Co-expression Group Robustness in Thaloid Haploid Tissues

To examine co-expression group robustness and mitigate the biasing effects elicited by the non-thaloid sporelings, antheridia, and diploid sporophytic tissues, we generated a co-expression matrix without these three groups (**Supplementary Figure S11**). In this matrix, MpARF1 forms a co-expression group with MpGRF, MpMADS2, MpCOI, and MIR11707, an amiR targeting MpAGO1 (Tsuzuki et al., 2015; Lin et al., 2016), that remains correlated with the auxin group. As expected, the auxin group now includes MpTAA and MpNCIAA but excludes MpNCARF and MpAUX/LAX. The biggest co-expression groups in this matrix are formed by reproductive genes or antheridiophore-specific genes (**Supplementary Figure S11**). Within the reproductive group, former groups 2 and 7 have merged to include MpARF3, MpSPL1-3, MpC3HDZ, MpCAMTA, MpNCARF, and MpTPL, which was previously associated with antheridia. MpAUX/LAX is grouped with the light signaling group, which has gained MpLUX, formerly associated with sporophytes. MpWIP, MpKAN, and MpRR-A form their own group, which has overlap with the auxin group.

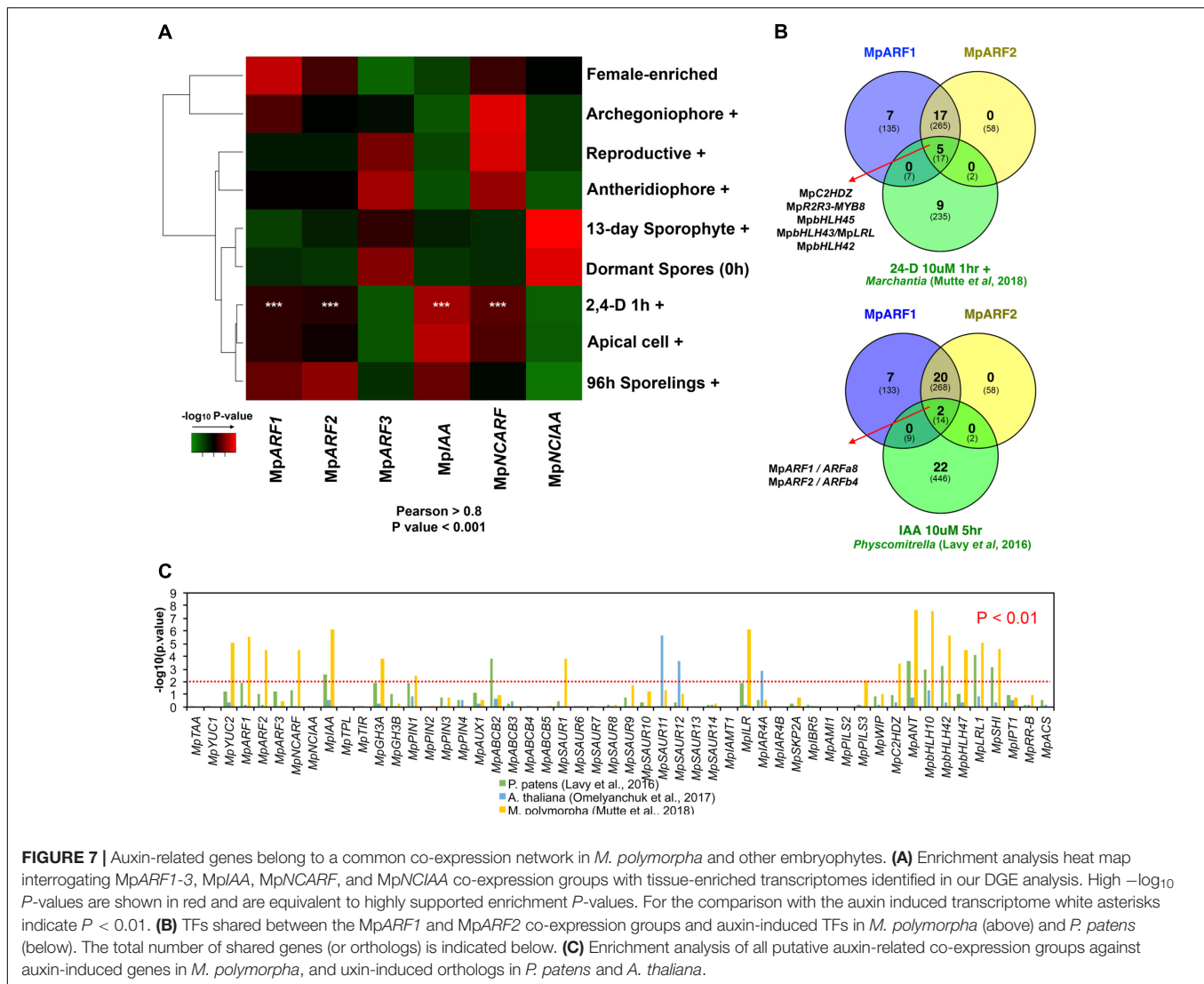
Co-expression Groups Are Dynamic at the Earliest Stages of Development

To examine how co-expression groups form at the earliest developmental transitions, we generated a matrix exclusive to sporeling RNA libraries. Multiple sporeling co-expression groups are formed that have different affinities with general co-expression groups (**Supplementary Figures S2, S3, S5, S12**). Auxin-related genes in particular show independent expression patterns prior to their clustering in later developmental stages. For example, the auxin signaling repressor MpIAA forms a well-supported group (Pearson > 0.9) with genes that putatively promote cell proliferation such as the class I TCP gene MpTCP1 (Martín-Trillo and Cubas, 2010; Davière et al., 2014), and the day 1 sporeling-enriched TF MpGRF (**Supplementary Figure S2**).

Additional genes in this group include MpC4HDZ, MpBBX3 and 5, MpGRAS8, MpDOF1 (Pearson > 0.9). MpGRAS4 is also correlated with GRF/TCP1, although more distantly (Pearson > 0.8). In sporelings, MpARF1 and MpARF2 form a well-supported group (Pearson = 0.93) with MpARF1 more closely sharing expression with MpMIR11707/MIRAGO1 (Pearson = 0.99), similar to what is observed in the thallus. MpRAV, MpbHLH48, MpANT, MpR2R3-MYB17, MpbHLH10, 21, 23 MpWIP and MpLOG are all co-expressed genes in this group (Pearson > 0.9). MpYUC2, MpNCIAA, and MpTIR form separate groups from each other and independent from the MpARF1/2 group. With MpTIR and MpNCIAA moderately close to each other (Pearson > 0.75). The largest co-expression group in sporelings consist of genes with high expression in 0-h sporelings. Genes included in this group are MpARF3, MpTAA, MpPIN1, MpSPL2, MpWOX, MpC3HDZ, MpVAL among others. Finally MpNCARF is in another group clustering with MpHMGbox3, MpRR-MYB6, MpHD2/SAWADEE, MpPIN4, and MpIPT1 (Pearson > 0.9).

The Auxin Co-expression Group Is Expressed Independently of the Single *Marchantia* Class C ARF and NCIAA

To validate a putative auxin co-expression group we performed a co-expression analysis of most PB1-containing genes (MpARF1-3, MpIAA, MpNCARF, and MpNCIAA) against all annotated *M. polymorpha* genes (**Figure 7**). Enrichment analysis (Fisher's exact test) was used to compare co-expression groups (Pearson > 0.8 and $P < 0.001$) with the multiple datasets obtained from our DGE analysis (**Figure 2**, $P < 0.01$, logFC > 0). The MpARF1, MpARF2, MpIAA, and MpNCARF co-expression groups were significantly composed of auxin-induced genes (Mutte et al., 2018) as well as apical cell-enriched transcripts (**Figure 7A**). Consistent with the TF co-expression matrix, the MpARF3 and MpNCIAA groups are not enriched in auxin-induced transcripts. TFs shared between the MpARF1 and MpARF2 groups and the Mutte et al. (2018) auxin-treated upregulated genes (**Figure 7B**) are MpC2HDZ, MpR2R3-MYB8, MpbHLH45, MpbHLH43/MpLRL, and MpbHLH42. The MpARF1, MpARF2, and MpIAA groups were also enriched in 96-h sporeling transcripts, particularly the MpARF2 group (**Figure 7A**). All groups had significant overlap with archegoniophore transcripts, except MpIAA. Only the MpARF3 and MpNCARF groups were enriched with reproductive transition genes (**Figure 7A**). Finally, the MpNCIAA group was significantly composed of dormant spores (**Supplementary Table S3**) and 13-day sporophyte enriched genes (**Figure 7A**). UpSet diagrams (**Supplementary Figure S13**) corroborated these observations with MpNCIAA having the largest unique co-expression group (588 genes exclusive to NCIAA), followed by MpARF3 ($N = 138$), MpNCARF ($N = 120$), MpIAA ($N = 83$), MpARF1 ($N = 58$), and finally MpARF2 ($N = 14$). The most correlated groups were MpARF1, MpARF2, MpNCARF, and MpIAA with 123 genes exclusively shared between them, followed by MpARF1, MpARF2, and MpIAA (66 exclusively shared genes) and MpARF1, MpARF2, and MpNCARF (51



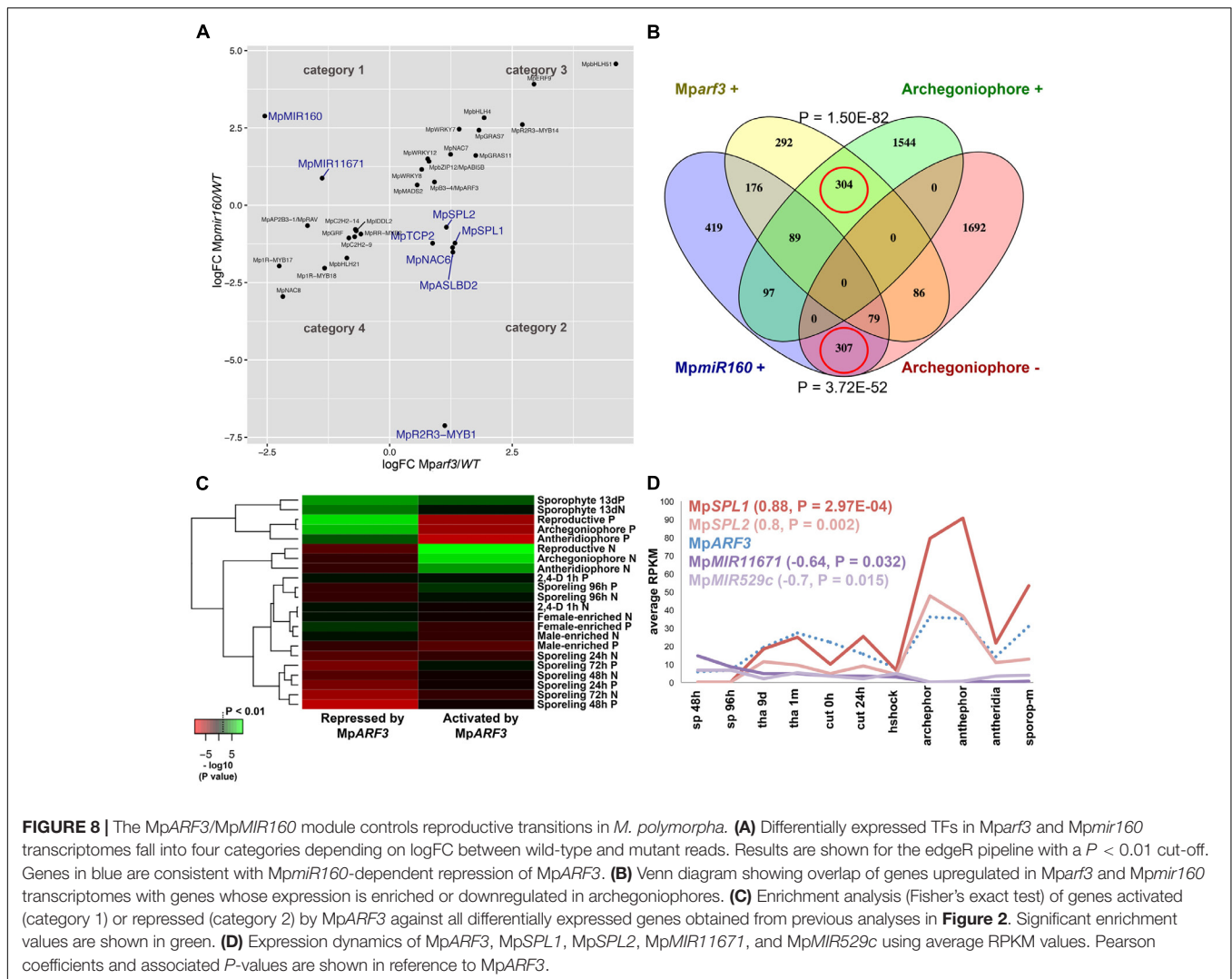
exclusively shared genes). No gene was shared across all groups, with only 21 genes shared between MpARF1-3, MpNCARF and MpIAA. Finally, MpARF3 and MpNCARF co-expression groups shared exclusively 25 genes (Figure 7A).

Consistent with these results, the MpARF1, MpARF2, MpNCARF, and MpIAA groups share six GO-terms by biological process (Supplementary Figure S14A) and four GO-terms (biological process) are shared between MpARF1, MpARF2 and MpIAA. In contrast, the MpARF3 group showed 12 specific GO-terms (Supplementary Figure S14B) while the MpNCIAA group had 11 specific GO-terms by biological process (Supplementary Figure S14C).

The Auxin Co-expression Group Is Conserved in Bryophytes

To further test the robustness of auxin-related co-expression groups, we compared them with differentially expressed genes (up and down-regulated) in auxin-treated *M. polymorpha*

thalli (Mutte et al., 2018), *P. patens* protonemal tissue (Lavy et al., 2016) and *Arabidopsis* root tissue (Omelyanchuk et al., 2017) using enrichment analysis. The co-expression groups of MpARF1, MpARF2, MpNCARF, MpIAA, MpYUC2, MpGH3A, MpPIN1, MpSAUR1, MpILR1, MpPILS3, MpC2HDZ, MpANT, MpHLLH10, MpHLLH42, MpHLLH47, MpHLLH43/MpLRL, and MpSHI showed significant similarity ($P < 0.01$, Fisher's exact-test) with DEGs in auxin-treated *Marchantia* plants (Figure 7C). Meanwhile, the MpIAA, MpABC2, MpANT, MpHLLH10, MpHLLH42, MpHLLH47, MpHLLH43/MpLRL, and MpSHI groups showed significant resemblance ($P < 0.01$) to *Physcomitrella* auxin treated upregulation profiles, with MpARF1, PIN1, MpGH3A, and MpILR groups overlapping at $P < 0.05$ (Figure 7C). In contrast, only the MpSAUR11, MpSAUR12, and MpIARF4A groups showed significant enrichment with auxin-treated *Arabidopsis* plants orthologs. This suggests a conservation of auxin co-expression groups across bryophytes, although the only TFs shared between the MpARF1 and MpARF2 co-expression groups

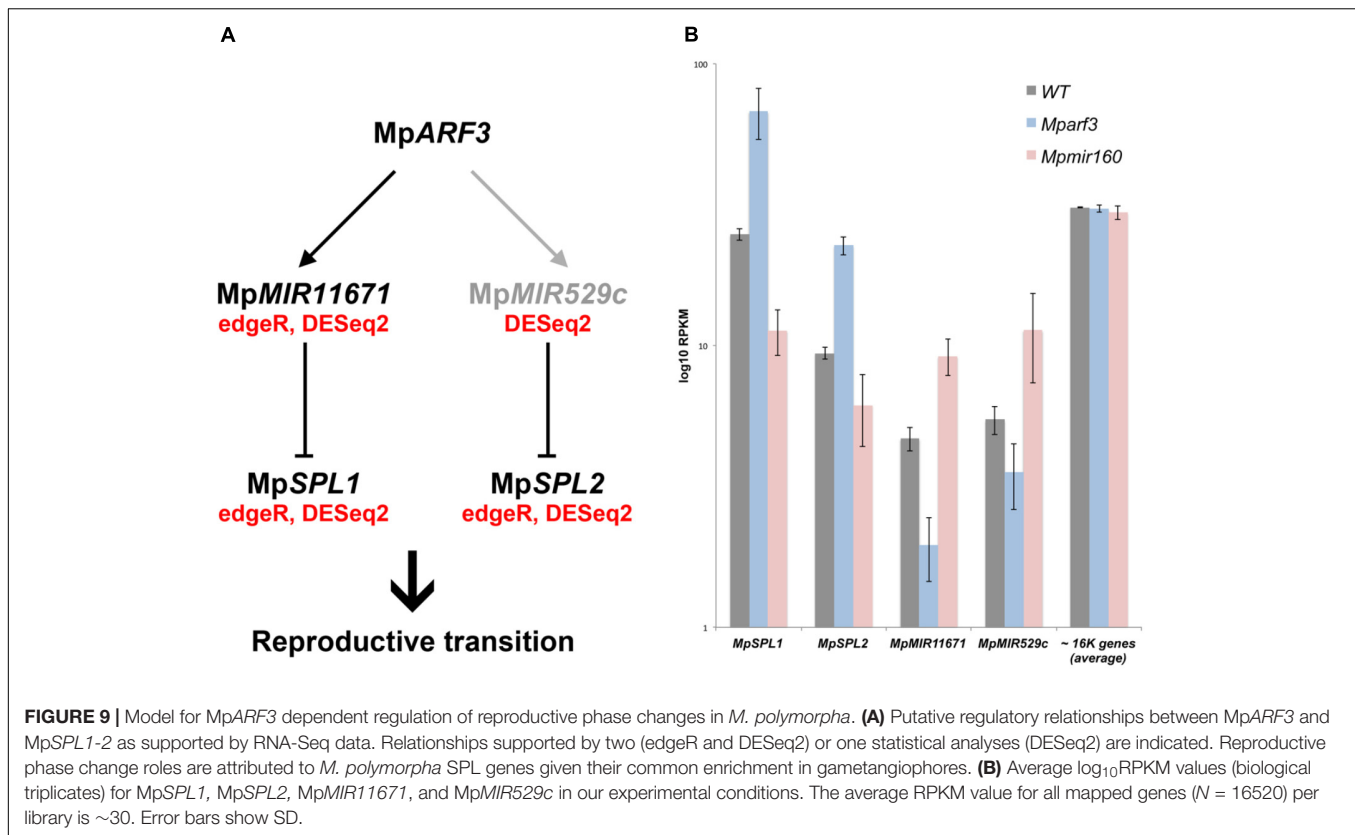


and the moss auxin-induced transcriptome are the class A and B ARF orthologs themselves, ARFa8 and ARFb4 (**Figure 7B**).

Analysis of *MpARF3* Gain and Loss-of-Function Mutant Transcriptomes Confirms Validity of Co-expression Groups

We performed DGE analysis (edgeR and DESeq2) in previously isolated *Mparf3/Mpmir160* CRISPR mutants (Flores-Sandoval et al., 2018) to test whether: (1) loss of *MpARF3* disrupts the auxin co-expression group, (2) genes co-expressed with *MpARF3* are altered in mutant alleles, and (3) *Mparf3* or *Mpmir160* transcriptomes could facilitate identification of genes responsible for developmental aspects of their respective mutant phenotypes. The edgeR analysis identified 2874 differentially expressed genes in *Mparf3* compared to the wild type ($P < 0.01$, 1524 upregulated). Similarly, 2864 differentially expressed genes were identified in comparisons of *Mpmir160* mutants with

the wild type ($P < 0.01$, 1572 upregulated). To filter out steady-state false positives and highlight genes dependent on *MpARF3* activity, we searched for genes with converse behaviors in the *Mparf3* and *Mpmir160* transcriptomes. Genes down-regulated in *Mparf3* and upregulated in *Mpmir160* mutants were considered to be activated, either directly or indirectly, by *MpARF3* (Category 1, $N = 112$; **Supplementary Table S18**). Conversely, genes upregulated in *Mparf3* and down-regulated in *Mpmir160* mutants were considered repressed, either directly or indirectly, by *MpARF3* (Category 2, $N = 115$; **Supplementary Table S19**). Strikingly, genes with similar dynamics, i.e., upregulated (Category 3, $N = 501$; **Supplementary Table S20**) or down-regulated (Category 4, $N = 470$; **Supplementary Table S21**) in both mutants were $\sim 4\times$ more abundant, suggesting a large cohort of genes affected by secondary feedback effects independent of the *Mpmir160/MpARF3* regulatory module (**Figure 8A**). The most prevalent protein family found in Category 1 was peroxidases (PF00141, **Supplementary Table S22**), while Category 2 had an abundance of FB_lectins (PF07367, **Supplementary Table S23**), four of which were in



tandem arrays in scaffold 118 and with high resemblance to fungal lectin genes (Bowman et al., 2017).

Although no TFs were identified as activated by MpARF3 (Category 1), two miR precursors fell into this category, with MpMIR160 (Mapoly0002s0211) and MpMIR11671 (Mapoly0239s0004) supported by both edgeR and DESeq2 analyses (Figure 8A). Meanwhile, a third miR precursor, MpMIR529c (Mapoly0006s0020) was also identified as activated by MpARF3 using DESeq2 (Supplementary Figure S14D). We had previously suggested that MpARF3 forms a negative feedback loop with MpmiR160 to control both developmental transitions and differentiation vs. totipotency in a context-dependent fashion (Flores-Sandoval et al., 2018).

The other two miRs identified in our analysis target SPL transcripts (Tsuzuki et al., 2015; Lin et al., 2016), with MpmiR11671 targeting MpSPL1 (Mapoly0014s0224, orthologous to SPL8 of *Arabidopsis*) and MpmiR529c targeting MpSPL2 (Mapoly0014s0223, orthologous to SPL3-5 of *Arabidopsis*). Consistent with these results, both MpSPL1 and MpSPL2 appeared as category 2 genes (i.e., repressed by MpARF3) using both edgeR and DESeq2, and supported by protein family GO-terms ($P = 0.0006924$). Additional TFs identified as downregulated by MpARF3 included MpTCP2 (Mapoly0001s0298), a FEZ-like NAC domain ortholog MpNAC6 (Mapoly0175s0015), the class Ia ASL1 gene MpASLBD2 (Mapoly0008s0060), and MpR2R3-MYB1 (Mapoly0001s0061). MpR2R3-MYB1 was most dramatically downregulated in Mpmir160 alleles (logFC = -7, Figure 8A).

Three of the downregulated TFs (MpSPL1, 2 and MpR2R3-MYB1) as well as MpARF3 itself were classified as reproductive genes in our wild-type DGE analysis (Figure 5). Furthermore, we had previously characterized MpARF3 as an antagonist of reproductive transitions, with Mpmir160 alleles insensitive to gametangiophore-inducing far-red light treatment and Mparf3 alleles as hypersensitive to far-red light treatment (forming more gametangiophores per area than wild type; Flores-Sandoval et al., 2018). We therefore measured the overlap between genes enriched in gametangiophores (either male or female) and the Mparf3 or Mpmir160 mutants. A significant number of genes ($N = 304$, $P = 1.5 \times 10^{-82}$, Hypergeometric test) were upregulated in both Mparf3 mutants and wild-type archegoniophores (Figure 8B). Consistently, a significant number of genes were upregulated in Mpmir160 mutants and downregulated in archegoniophores ($N = 307$, $P = 3.7 \times 10^{-52}$, Hypergeometric test). Enrichment analysis (Fisher's exact test) using the datasets obtained in our DGE analysis (edgeR, $P < 0.01$) show that overall transcripts activated by MpARF3 (category 1) are similar to those downregulated in wild type gametangiophores ($P < 0.01$, Figure 8C). Conversely, transcripts repressed by MpARF3 (category 2) are significantly upregulated in wild type gametangiophores, in particular the subset of reproductive transition genes and not the female or male-enriched genes (Figure 8C). Additionally, 13-day sporophytes also have transcripts that seem to be controlled by MpARF3 although with an ambiguous trend unlike that observed for reproductive tissues. Although some auxin-related

or auxin-induced genes were differentially expressed in *Mparf3* transcriptomes (Flores-Sandoval et al., 2018), not a single auxin-induced gene had converse expression patterns in the *Mparf3* and *Mpmir160* transcriptomes (Figure 8A). Finally, the expression dynamics of the candidate TF/MIR pairs shows that *MpARF3*, *MpSPL1* and 2 expression patterns are positively correlated with peaks of expression in gametangiophores (Figure 8D), while the *MpMIR11671* and *MpMIR529c* precursors are negatively correlated with *MpARF3* and show down-regulation in gametangiophores. Thus, an inverse relationship exists between *MpARF3* activity and expression of the *MpSPL1-2/MpMIR529c/MpMIR11671* modules in *M. polymorpha* (Figures 8D, 9B).

DISCUSSION

This study provides a first attempt at characterizing DGE patterns and co-expression groups throughout the life cycle of *M. polymorpha* (Figure 1). Given their hierarchical regulatory character, TFs were chosen as a proxy for the thousands of genes differentially expressed between cell and tissue types. DGE analysis reveals that only a minority of TFs show dramatic fold-changes between developmental transitions in *M. polymorpha*. However, this figure is qualified by the limited number of tissues and environmental conditions sampled in our analysis (Figure 2). These differentially expressed TFs are candidates for functional analysis to assess whether they determine cell identity or broader processes required for tissue patterning, i.e., meiosis, cell division, expansion, stress-responses.

The relatively small number of annotated TFs (~400) in *M. polymorpha* (Bowman et al., 2017) facilitated construction of a co-expression matrix using 11 tissues, including sporelings, thallus, gametangiophores, antheridia and sporophytes. This matrix resolved multiple discrete groups with variable levels of overlap and putative interactions (Figure 6). As expected, some groups correspond to tissue-specific TF enrichment (e.g., sporophytes, antheridia, archegoniophores or regenerating tissue). Surprisingly, a second kind of group that did not show enrichment in any one type of library (Supplementary Figures S9E,G,I,L) was also identified. These groups include the partners of *MpARF1* or *MpTOC1*, which are involved in auxin and light/circadian responses, respectively. Members of these co-expression groups were differentially expressed in multiple stages but did not change dramatically across our pair-wise comparisons, likely due to lack of inductive environmental conditions or because they are under tight regulatory feedbacks. Thus, our data suggests developmental transitions are also defined by the additive influence of multiple members of a co-expression group.

TF Functions Predicted by Our Analysis

Totipotency

A set of TFs that could represent regulating factors controlling aspects of totipotency include the *LAS* ortholog *MpGRAS4*, *MpANT*, and *MpARF2*. *MpGRAS4* is enriched in the apical cell and both 24 and 48 h sporelings. It is further nested

in a co-expression group enriched in 24-h regenerating tissue following wounding (Supplementary Figure S9F). Thus, *MpGRAS4* is expressed in tissues united by active cell proliferation. Second, the *ANT/PLT/BBM* ortholog, *MpANT*, is enriched exclusively in apical cells and 96 h sporelings, raising the possibility that its action is downstream of *MpGRAS4*. *MpANT* is also co-expressed in the auxin cluster (Supplementary Figure S10), suggesting that it may be auxin-inducible, as occurs in angiosperm model systems (Galinha et al., 2007; Ding and Friml, 2010). Consistent with this scenario, *MpARF2* is also an apical-cell enriched gene whose expression steadily increases from 72-h sporelings (in parallel with *MpARF1*, Supplementary Figure S4A). It is plausible, given that *MpARF1* and *MpARF2* act as respective activators or repressors of transcription (Kato et al., 2015), that antagonism between *MpARF1* and *MpARF2* at 72 h stabilizes *MpANT* expression in 96 h sporelings, subsequently forming a co-expression group throughout development. Given that *LAS* has key roles specifying axillary meristems in angiosperms, and that *ANT/PLT/BBM* orthologs act in meristems in both angiosperms (Aida et al., 2004) and mosses (Aoyama et al., 2012), there might be more overlap between haploid and diploid meristems than previously acknowledged (Frank and Scanlon, 2015).

Cell Proliferation and Auxin

Additional genes may play a role in cellular processes that contribute to, but do not specify, meristematic tissues. For example, a cluster represented by the 24-h sporelings involving *MpTCP1*, *MpGRF*, *MpC4HDZ* and *MpIAA* genes could represent a cell proliferation group. *MpGRF* in particular is highly expressed in the first day of sporeling germination, but it is also co-expressed with auxin-related genes throughout the life cycle (Figure 6A and Supplementary Figure S11). Auxin signaling in 24-h sporelings may be constrained by the presence of the known *MpIAA* auxin signaling repressor (Flores-Sandoval et al., 2015; Kato et al., 2015). Given that key aspects of auxin physiology in *M. polymorpha* involve organ differentiation (Eklund et al., 2015; Flores-Sandoval et al., 2015; Kato et al., 2015), it is tempting to speculate that there may be an antagonism between auxin-dependent differentiation and cell proliferation elicited by *MpGRF* and other TFs in *M. polymorpha*.

Differentiation Factors

An auxin-inducible gene (Mutte et al., 2018) characterized in *M. polymorpha* is the IDDL-like Zinc Finger *MpWIP* TF, which is essential for air pore differentiation (Jones and Dolan, 2017). Consistently, *MpWIP* is upregulated in 96-h sporelings at the time of photosynthetic tissue proliferation (Supplementary Figure S5). *MpWIP* is co-expressed (Supplementary Figure S9H) with the GARP TF *MpKAN*, whose orthologs influence organ polarity in angiosperms (Eshed et al., 2001). *MpKAN* and its co-expression partner *MpWRKY8* are in turn clearly enriched in 0-h cut thalli that lack apical notches and haven't initiated regeneration. One hypothesis is that members of this group may regulate key aspects of differentiation possibly downstream of auxin and light signaling, given their overlap in expression patterns (*MpKAN* and *MpARF1* have a Pearson

coefficient of 0.76, while MpPIF and MpKAN have a coefficient of 0.8).

Reproductive Genes in *Marchantia*

Comparisons between gametangiophore datasets facilitated distinctions between reproductive transition TFs (enriched in both sexes) vs. female or male-enriched TFs. MpBNB served as a marker to validate this group given its role in promoting formation of gametangiophores irrespective of sex (Yamaoka et al., 2018). A co-expression group including MpSPL1, MpSPL2, MpC3HDZIP, MpABI3B, MpARF3, among others correlates with the vegetative to reproductive transition and is supported by mutant transcriptomes (see below section). Interestingly, auxin-repressed genes are significantly represented in archegoniophore and antheridiophore-enriched transcripts (Supplementary Figure S4B), suggesting a repressor(s) of auxin signaling is a member of the reproductive transition group. Male-enriched TFs were refined by the presence of antheridia libraries, which suggest a robust set of male gamete-enriched TFs and these have been extensively described in previous studies (Higo et al., 2016). MpRWP2/MpMID could provide an interesting candidate for an antheridia specific factor given that its orthologs specify *minus* gametes in green algae (Ferris and Goodenough, 1997).

Extended Parental to Zygotic Transition in *M. polymorpha*

The identification of female and antheridia-enriched TFs allowed comparisons with upregulated TFs in 13-day sporophytes, consisting of ~10 cells and at which point sporogenous tissue differentiation has not commenced (Frank and Scanlon, 2015). Female-enriched TFs shared with young sporophytes include members of 10 TF families. Male-enriched TFs shared with young sporophytes involve members of 7 TF families, although MpBELL4 is the only identified TF exclusively upregulated in antheridia and young sporophytes (Figure 2B). Enrichment analysis (Supplementary Figure S6) suggests that only archegoniophore-enriched genes are significantly represented in 13-day sporophyte transcriptomes. Thus, the possibility of RNA contamination from maternal tissues still remains as a plausible explanation to this phenomenon.

Class C ARF and MpNCIAA Genes Are Independent of Other Auxin Related Genes

The discovery of a putative auxin co-expression group among transcription factors (Figure 6) coincides with a similar PIN1-related co-expression group observed in moss (Ruprecht et al., 2017b), suggesting an ancestral origin for the auxin co-expression network in land plants. The robustness of the auxin TF co-expression group was tested using whole-genome co-expression analysis for candidate genes and comparisons with the only available auxin-inducible transcriptome in *M. polymorpha* (Mutte et al., 2018). Enrichment analysis supports significant overlap between MpARF1, MpARF2, MpIAA, MpNCARF, MpYUC2, MpSAUR1, MpPIN1, MpGH3A, and MpILR1 co-expression groups and the *M. polymorpha* auxin-upregulated

transcriptome (Figures 7A,C). Other TFs within the auxin co-expression group but strictly not involved in auxin biosynthesis, transport, conjugation or responses include MpANT, MpSHI, MpbHLH10, 43, 45 and 47, MpMIR11707, MpR2R3-MYB8, and MpC2HDZ (Figure 7C and Supplementary Figure S12). They could represent upstream or downstream elements regulating the module (Mutte et al., 2018) and are suitable candidates for future characterization. Importantly, many of these genes are also part of the auxin-inducible transcriptome of *Physcomitrella* but not *Arabidopsis* (Figure 7C), suggesting they were key components of the auxin co-expression network in the ancient bryophyte ancestor. As MpbHLH43/LRL possibly mediates ectopic rhizoid formation (Breuninger et al., 2016) in response to exogenous auxin in *M. polymorpha* its association was anticipated. Furthermore, 5 of the 14 TFs induced by auxin (Supplementary Table S24) in *M. polymorpha* are found within the MpARF1 and MpARF2 co-expression groups (Figure 7B). Notably, the class C ARF, MpARF3, a gene that is not necessary for physiological and transcriptional auxin responses (Flores-Sandoval et al., 2018; Mutte et al., 2018), forms part of an independent co-expression group highly active in reproductive structures (Figure 8 and Supplementary Figure S9). However, MpARF3 still shares members with MpNCARF (72 genes), MpARF1/MpARF2/MpNCARF/MpIAA (21 genes) and MpNCIAA (9 genes, Supplementary Figure S13) co-expression groups. The significance of this connectivity was tested (Supplementary Table S25), providing ways in which MpARF3 competes for MpARF1 and MpARF2 targets in an auxin-independent manner. Importantly, the MpARF3 and MpNCARF co-expression groups are enriched in reproductive transition genes (Figure 7A), opening the possibility of interaction at this developmental stage. MpNCIAA is another PBI-containing factor that forms an independent group with sporophyte-enriched transcripts (Figure 7 and Supplementary Figure S9) and is also not essential for transcriptional auxin responses (Mutte et al., 2018). Consistent with their auxin independence, both class C ARFs and NCIAAs evolved before the origin the canonical auxin-signaling pathway (and land plants), and thus they may act in independent regulatory networks. The putative roles of NCIAA genes in patterning sporophytic development are of interest although these genes are also expressed in haploid tissues in *M. polymorpha*.

RNA-Seq Data Supports the Role of MpARF3 as an Inhibitor of Reproductive Transitions

We have previously described MpARF3 as a repressor of differentiation in multiple developmental contexts. For example, MpARF3 mutants display developmental transition, defects, such as ectopic differentiation of air pores, scales, pegged rhizoids, and gametophores, and are deficient in forming undifferentiated gemmalings (Flores-Sandoval et al., 2018; Mutte et al., 2018). Meanwhile strong MpARF3 gain-of-function alleles form undifferentiated calli whose cell identity resembles that of young sporelings or prothalli (Flores-Sandoval et al., 2018). Thus, we anticipated the transcriptome data should reflect

aspects of these developmental defects. Using gain- and loss-of-function MpARF3 transcriptomes, we were able to discern secondary feedback, treatment-dependent and steady state effects and thus assign with more confidence genes genuinely dependent on the MpMIR160/MpARF3 regulatory module (category 1 and 2 genes, **Figure 8**). Indeed, enrichment analysis demonstrated that MpARF3 represses, directly or indirectly, genes that promote reproductive transitions (**Figure 8**). Consistently, MpARF3 activates genes that inhibit production of, or at least are downregulated in, sexual organs. We therefore propose the hypothesis (**Figure 9A**) that MpARF3 inhibits the reproductive transition in *M. polymorpha* via activation of the MpMIR11671 and MpMIR529c precursors, which in turn target MpSPL1 and MpSPL2. This appears plausible, given the observed MpSPL1 and MpSPL2 RPKM values in Mp*arf3*/mir160 mutant backgrounds (**Figure 9B**) and the roles of SPL genes in regulating heteroblastic changes and specifically in promoting reproductive transitions in angiosperms (Huijser and Schmid, 2011) and possibly mosses (Cho et al., 2012).

AUTHOR CONTRIBUTIONS

EF-S generated data for **Figures 1–5**. EF-S and FR generated data for **Figures 6–8**. FR created scripts for enrichment analysis and GO-terms pipelines. EF-S designed all figures, except **Figures 6, 7**, which were jointly designed by EF-S and FR. EF-S and JB co-wrote the manuscript and interpreted the data. All authors analyzed and revised the data.

FUNDING

This work was made possible by funding from the Australian Research Council (DP160100892 and DP170100049 to JB) and by the Agencia Nacional de Promoción Científica y Tecnológica (PICT2013-3285 and PICT2017-1484).

ACKNOWLEDGMENTS

We thank all past and current members of the Bowman Lab for feedback and helpful discussions. The Monash University Bioinformatics platform provided helpful guidance and feedback. The Monash Next Generation sequencing facility (Micromon) sequenced the Mp*arf3* and Mp*mir160* transcriptomes. We thank Sandra K. Floyd, Tom Dierschke, and John Alvarez for providing images for **Figure 1**.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.01345/full#supplementary-material>

FIGURE S1 | (A) Frequency of log2FC values in pairwise comparisons performed in this study. **(B)** UpSet diagram of upregulated genes (logFC), including all

curated TFs, in this study. Red asterisk indicates TFs not differentially expressed in any DGE analysis comparison.

FIGURE S2 | Identification of co-expression groups of 24-h enriched TFs in *M. polymorpha* sporplings. RPKM values are shown for all replicates from 0 to 96 h.

FIGURE S3 | Identification of co-expression groups of 48-h enriched TFs *M. polymorpha* sporplings. RPKM values are shown for all replicates from 0 to 96 h.

FIGURE S4 | Auxin-signaling and biosynthetic gene expression during sporpling development. **(A)** RPKM averages of MpARF1, MpARF2, MpARF3, MpIAA, MpTAA, and MpYUC2 during the first four days of sporpling development. Error bars indicate SD. Upregulation of MpARF1 and MpARF2 and downregulation of MpIAA at 96 h are statistically supported. *P*-values marked are from edgeR analysis. **(B)** Enrichment analysis of auxin upregulated and downregulated genes probed against multiple developmental RNA-Seq libraries obtained from DGE analysis. The tissue with the highest resemblance to the auxin-upregulated transcriptome is 96-h sporplings.

FIGURE S5 | Identification of co-expression groups of 72 and 96-h enriched TFs *M. polymorpha* sporplings. RPKM values are shown for all replicates from 0 to 96 h.

FIGURE S6 | Sporophytic gene expression in *M. polymorpha*. **(A)** Enrichment analysis of 13-day and JGI manually dissected sporophytes shows significant enrichment between the transcriptomes. Genes classified as female-enriched (**Supplementary Table S10**) are also statistically represented in 13-day sporophyte transcriptomes. **(B)** Venn Diagram of shared TFs between antheridia, archegoniophore, 13-day and JGI sporophytes. Putative maternal, paternal and common TFs continuing expression in 13-day sporophytes are annotated. The + symbol indicates upregulation (logFC > 0).

FIGURE S7 | UpSet diagram of tissue-specific TF enrichment at logFC > 1. Red asterisk shows eight uniquely upregulated TFs in apical cells at logFC > 1 compared to other tissues.

FIGURE S8 | Distance dendrogram generated with hclust (R Version 2.3.2) of all annotated *M. polymorpha* TFs (**Supplementary Table S1**) in 11 tissue libraries using RPKM > 1 as a threshold.

FIGURE S9 | RPKM dynamics of TF/response/biosynthetic genes obtained in our analysis. Pearson coefficients above 0.8 were used to define the MpBELL4 **(A)**, MpARF3 **(B)**, MpBELL1 **(C)**, MpNCIAA **(D)**, MpBHLH14/RSL **(E)**, MpGRAS4 **(F)**, MpTOC1 **(G)**, MpGARP1/KAN **(H)**, MpGARP8/GLD **(I)**, MpSPL2 **(J)**, MpNAC8 **(K)**, and MpARF1 **(L)** coexpression groups. Accompanying TFs are indicated at the left of each graph.

FIGURE S10 | RPKM dynamics of the putative auxin co-expression group genes.

FIGURE S11 | (A) Heatmap of Pearson coefficient matrix of all annotated *M. polymorpha* TFs, MIR precursors and putative hormonal genes (**Supplementary Table S16**) RPKM values above 0.75 (to include MpANT) in sporpling tissues. Putative groups were delimited by sets of genes sharing coefficients above 0.8. Genes putatively involved in auxin biosynthesis, perception, signaling and transport are mapped, together with accompanying partners in smaller fonts. **(B)** Distance dendrogram of tissue libraries used in the analysis. Average replicate RPKM values were used per library. Scale for heatmap indicates Pearson coefficients.

FIGURE S12 | (A) Heatmap of Pearson coefficient matrix of all annotated *M. polymorpha* TFs, MIR precursors and putative hormonal genes with RPKM values above 5 (**Supplementary Table S17**) in thalli and gametangioophores. Putative groups were delimited by sets of genes sharing coefficients above 0.8. Genes putatively involved in auxin biosynthesis, perception, signaling and transport are mapped, together with accompanying partners in smaller fonts. **(B)** Distance dendrogram of tissue libraries used in the analysis. Average replicate RPKM values were used per library. Scale for heatmap indicates Pearson coefficients.

FIGURE S13 | (A) UpSet diagram showing shared co-expressed genes (cut-offs indicated) between most PBI-containing genes in *M. polymorpha*.

Numbers of genes exclusive or shared between groups are shown. Co-expression analyses were performed for all expressed genes in our RPKM matrix.

FIGURE S14 | (A) Significant GO-terms (Biological Process) found in the MpARF1, MpARF2, MpNCARF, and MpIAA co-expression groups. X-axis indicates *P*-Values. **(B)** Significant GO-terms (Biological Process) found in the MpARF3 co-expression group. X-axis indicates *P*-values. **(C)** Significant GO-terms (Biological Process) found in the MpNCIAA co-expression group. X-axis indicates *P*-values. **(D)** TFs differentially expressed in Mparf3 and Mpmir160 transcriptomes using DESeq2 ($P < 0.01$ cut-off). Genes in blue are consistent with Mpmir160-dependent repression of MpARF3.

TABLE S1 | TFs, MIRs and hormonal genes and accession numbers used in this paper.

TABLE S2 | Shared upregulated genes between DEG analysis comparisons at logFC > 0.

TABLE S3 | Differentially expressed TFs and genes in 24-h sporelings using edgeR.

TABLE S4 | Differentially expressed TFs and genes in 48-h sporelings using edgeR.

TABLE S5 | Differentially expressed TFs and genes in 72-h sporelings using edgeR.

TABLE S6 | Differentially expressed TFs and genes in 96-h sporelings using edgeR.

TABLE S7 | Differentially expressed TFs and genes in archegoniophores using edgeR.

TABLE S8 | Differentially expressed TFs and genes in antheridiophores using edgeR.

TABLE S9 | Differentially expressed TFs and genes shared in archegoniophores and antheridiophores (reproductive TFs) using edgeR.

TABLE S10 | Archegoniophore-enriched TFs and genes using edgeR.

TABLE S11 | Antheridiophore-enriched TFs and genes using edgeR.

TABLE S12 | Differentially expressed TFs and genes in antheridia using edgeR.

TABLE S13 | Differentially expressed TFs and genes in 13-day sporophytes using edgeR.

TABLE S14 | Shared upregulated genes between DEG analysis comparisons at logFC > 1.

TABLE S15 | Pearson coefficient matrix of TFs for **Figure 6**.

TABLE S16 | Pearson coefficient matrix of TFs for **Supplementary Figure S11**.

TABLE S17 | Pearson coefficient matrix of TFs for **Supplementary Figure S12**.

TABLE S18 | Category 1 genes from Mparf3 (x) and Mpmir160 (y) transcriptomes.

TABLE S19 | Category 2 genes from Mparf3 (x) and Mpmir160 (y) transcriptomes.

TABLE S20 | Category 3 genes from Mparf3 (x) and Mpmir160 (y) transcriptomes.

TABLE S21 | Category 4 genes from Mparf3 (x) and Mpmir160 (y) transcriptomes.

TABLE S22 | Go analysis (PF-terms) of Category 1 genes from Mparf3 (x) and Mpmir160 (y) transcriptomes.

TABLE S23 | Go analysis (PF-terms) of Category 2 genes from Mparf3 (x) and Mpmir160 (y) transcriptomes.

TABLE S24 | Differentially expressed TFs in auxin-induced tissue (1 h) from Mutte et al. (2018) using edgeR.

TABLE S25 | Enrichment-analysis between MpARF1-3, MpIAA, MpNCARF, and MpNCIAA co-expression groups, *P*-values are shown.

REFERENCES

- Afgan, E., Sloggett, C., Goonasekera, N., Makunin, I., Benson, D., Crowe, M., et al. (2015). Genomics virtual laboratory: a practical bioinformatics workbook for the cloud. *PLoS One* 10:e0140829. doi: 10.1371/journal.pone.0140829
- Aida, M., Beis, D., Heidstra, R., Willemsen, V., Blilou, I., Galinha, C., et al. (2004). The PLETHORA genes mediate patterning of the *Arabidopsis* root stem cell niche. *Cell* 119, 109–120. doi: 10.1016/j.cell.2004.09.018
- Alaba, S., Piszczalka, P., Pietrykowska, H., Pacak, A. M., Sierocka, I., Nuc, P. W., et al. (2015). The liverwort *Pellia endiviifolia* shares microtranscriptomic traits that are common to green algae and land plants. *New Phytol.* 206, 352–367. doi: 10.1111/nph.13220
- Aoyama, T., Hiwatashi, Y., Shigyo, M., Kofuji, R., Kubo, M., Ito, M., et al. (2012). AP2-type transcription factors determine stem cell identity in the moss *Physcomitrella patens*. *Development* 139, 3120–3129. doi: 10.1242/dev.076091
- Babicki, S., Arndt, D., Marcu, A., Liang, Y., Grant, J. R., Maciejewski, A., et al. (2016). Heatmapper: web-enabled heat mapping for all. *Nucleic Acids Res.* 44, W147–W153. doi: 10.1093/nar/gkw419
- Boehm, C. R., Pollak, B., Purswani, N., Patron, N., and Haseloff, J. (2017). Synthetic botany. *Cold Spring Harb. Perspect. Biol.* 9:a023887. doi: 10.1101/cshperspect.a023887
- 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.e15–304.e15. doi: 10.1016/j.cell.2017.09.030
- Breuninger, H., Thamm, A., Streubel, S., Sakayama, H., Nishiyama, T., and Dolan, L. (2016). diversification of a transcription factor family led to the evolution of antagonistically acting genetic regulators of root hair growth. *Curr. Biol.* 26, 1622–1628. doi: 10.1016/j.cub.2016.04.060
- Cassan-Wang, H., Goué, N., Saidi, M. N., Legay, S., Sivadon, P., Goffner, D., et al. (2013). Identification of novel transcription factors regulating secondary cell wall formation in *Arabidopsis*. *Front. Plant Sci.* 4:189. doi: 10.3389/fpls.2013.00189
- Chater, C. C., Caine, R. S., Tomek, M., Wallace, S., Kamisugi, Y., Cuming, A. C., et al. (2016). Origin and function of stomata in the moss *Physcomitrella patens*. *Nat. Plants* 2, 16179. doi: 10.1038/nplants.2016.179
- Cho, S. H., Coruh, C., and Axtell, M. J. (2012). miR156 and miR390 regulate tasiRNA accumulation and developmental timing in *Physcomitrella patens*. *Plant Cell* 24, 4837–4849. doi: 10.1105/tpc.112.103176
- Davière, J.-M., Wild, M., Regnault, T., Baumberger, N., Eisler, H., Genschik, P., et al. (2014). Class I TCP-DELLA interactions in inflorescence shoot apex determine plant height. *Curr. Biol.* 24, 1923–1928. doi: 10.1016/j.cub.2014.07.012
- Devos, N., Szövényi, P., Weston, D. J., Rothfels, C. J., Johnson, M. G., and Shaw, A. J. (2016). Analyses of transcriptome sequences reveal multiple ancient large-scale duplication events in the ancestor of Sphagnopsida (Bryophyta). *New Phytol.* 211, 300–318. doi: 10.1111/nph.13887
- Ding, Z., and Friml, J. (2010). Auxin regulates distal stem cell differentiation in *Arabidopsis* roots. *Proc. Natl. Acad. Sci. U.S.A.* 107, 12046–12051. doi: 10.1073/pnas.1000672107
- Eklund, D. M., Ishizaki, K., Flores-Sandoval, E., Kikuchi, S., Takebayashi, Y., Tsukamoto, S., et al. (2015). Auxin produced by the indole-3-pyruvic acid pathway regulates development and gemmae dormancy in the liverwort *Marchantia polymorpha*. *Plant Cell* 27, 1650–1669. doi: 10.1105/tpc.15.00065
- Eklund, D. M., Thelander, M., Landberg, K., Staldal, V., Nilsson, A., Johansson, M., et al. (2010). Homologues of the *Arabidopsis thaliana* SHI/STY-LRP1 genes control auxin biosynthesis and affect growth and development in the moss *Physcomitrella patens*. *Development* 137, 1275–1284. doi: 10.1242/Dev.039594
- Eshed, Y., Baum, S. F., Perea, J. V., and Bowman, J. L. (2001). Establishment of polarity in lateral organs of plants. *Curr. Biol.* 11, 1251–1260. doi: 10.1016/S0960-9822(01)00392-X
- Ferreira, S. S., Hotta, C. T., de Carli Poelking, V. G., Leite, D. C. C., Buckeridge, M. S., Loureiro, M. E., et al. (2016). Co-expression network analysis reveals

- transcription factors associated to cell wall biosynthesis in sugarcane. *Plant Mol. Biol.* 91, 15–35. doi: 10.1007/s11103-016-0434-2
- Ferris, P. J., and Goodenough, U. W. (1997). Mating type in *Chlamydomonas* is specified by mid, the minus-dominance gene. *Genetics* 146, 859–869.
- Ficklin, S. P., and Feltus, F. A. (2011). Gene coexpression network alignment and conservation of gene modules between two grass species: maize and rice. *Plant Physiol.* 156, 1244–1256. doi: 10.1104/pp.111.173047
- Flores-Sandoval, E., Dierschke, T., Fisher, T. J., and Bowman, J. L. (2016). Efficient and inducible use of artificial MicroRNAs in *Marchantia polymorpha*. *Plant Cell Physiol.* 57, 281–290. doi: 10.1093/pcp/pcv068
- 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
- Flores-Sandoval, E., Eklund, D. M., Hong, S. F., Alvarez, J. P., Fisher, T. J., Lampugnani, E. R., et al. (2018). Class C ARFs evolved before the origin of land plants and antagonize differentiation and developmental transitions in *Marchantia polymorpha*. *New Phytol.* 218, 1612–1630. doi: 10.1111/nph.15090
- Frank, M. H., and Scanlon, M. J. (2015). Transcriptomic evidence for the evolution of shoot meristem function in sporophyte-dominant land plants through concerted selection of ancestral gametophytic and sporophytic genetic programs. *Mol. Biol. Evol.* 32, 355–367. doi: 10.1093/molbev/msu303
- Galinha, C., Hofhuis, H., Luijten, M., Willemsen, V., Blilou, I., Heidstra, R., et al. (2007). PLETHORA proteins as dose-dependent master regulators of *Arabidopsis* root development. *Nature* 449, 1053–1057. doi: 10.1038/nature06206
- Gerstein, M. B., Rozowsky, J., Yan, K. K., Wang, D., Cheng, C., Brown, J. B., et al. (2014). Comparative analysis of the transcriptome across distant species. *Nature* 512, 445–448. doi: 10.1038/nature13424
- Greb, T., Clarenz, O., Schäfer, E., Müller, D., Herrero, R., Schmitz, G., et al. (2003). Molecular analysis of the LATERAL SUPPRESSOR gene in *Arabidopsis* reveals a conserved control mechanism for axillary meristem formation. *Genes Dev.* 17, 1175–1187. doi: 10.1101/gad.260703
- Higo, A., Niwa, M., Yamato, K. T., Yamada, L., Sawada, H., Sakamoto, T., et al. (2016). Transcriptional framework of male gametogenesis in the liverwort *Marchantia polymorpha* L. *Plant Cell Physiol.* 57, 325–338. doi: 10.1093/pcp/pcw005
- Hori, K., Maruyama, F., Fujisawa, T., Togashi, T., Yamamoto, N., Seo, M., et al. (2014). *Klebsormidium flaccidum* genome reveals primary factors for plant terrestrial adaptation. *Nat. Commun.* 5:3978. doi: 10.1038/ncomms4978
- Huijser, P., and Schmid, M. (2011). The control of developmental phase transitions in plants. *Development* 138, 4117–4129. doi: 10.1242/dev.063511
- Ichihashi, Y., Aguilar-Martínez, J. A., Farhi, M., Chitwood, D. H., Kumar, R., Millon, L. V., et al. (2014). Evolutionary developmental transcriptomics reveals a gene network module regulating interspecific diversity in plant leaf shape. *Proc. Natl. Acad. Sci. U.S.A.* 111, E2616–E2621. doi: 10.1073/pnas.1402835111
- Inoue, K., Nishihama, R., Kataoka, H., Hosaka, M., Manabe, R., Nomoto, M., et al. (2016). Phytochrome signaling is mediated by phytochrome interacting factor in the liverwort *Marchantia polymorpha*. *Plant Cell* 28, 1406–1421. doi: 10.1105/tpc.15.01063
- Ishizaki, K., Nishihama, R., Yamato, K. T., and Kohchi, T. (2016). Molecular genetic tools and techniques for *Marchantia polymorpha* research. *Plant Cell Physiol.* 57, 262–270. doi: 10.1093/pcp/pcv097
- Jones, V. A. S., and Dolan, L. (2017). Mp WIP regulates air pore complex development in the liverwort *Marchantia polymorpha*. *Development* 144, 1472–1476. doi: 10.1242/dev.144287
- Kareem, A., Durgaprasad, K., Sugimoto, K., Du, Y., Pulianmackal, A. J., Trivedi, Z. B., et al. (2015). PLETHORA genes control regeneration by a two-step mechanism. *Curr. Biol.* 25, 1017–1030. doi: 10.1016/j.cub.2015.02.022
- Kato, H., Ishizaki, K., Kouno, M., Shirakawa, M., Bowman, J. L., Nishihama, R., et al. (2015). Auxin-mediated transcriptional system with a minimal set of components is critical for morphogenesis through the life cycle in *Marchantia polymorpha*. *PLoS Genet.* 11:e1005084. doi: 10.1371/journal.pgen.1005084
- Kato, H., Kouno, M., Takeda, M., Suzuki, H., Ishizaki, K., Nishihama, R., et al. (2017). The roles of the sole activator-type auxin response factor in pattern formation of *Marchantia polymorpha*. *Plant Cell Physiol.* 58, 1642–1651. doi: 10.1093/pcp/pcx095
- Kemmeren, P., van Berkum, N. L., Vilo, J., Bijma, T., Donders, R., Brazma, A., et al. (2002). Protein interaction verification and functional annotation by integrated analysis of genome-scale data. *Mol. Cell* 9, 1133–1143. doi: 10.1016/S1097-2765(02)00531-2
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14:R36. doi: 10.1186/gb-2013-14-4-r36
- Kny, L. (1890). “Bau und entwicklung von *Marchantia polymorpha*,” in *Botanische Wandtafeln Mit Erläuterndem*, ed. V. von Wiegandt (Berlin: Hempel & Parey).
- Koi, S., Hisanaga, T., Sato, K., Shimamura, M., Yamato, K. T., Ishizaki, K., et al. (2016). An evolutionarily conserved plant RKD factor controls germ cell differentiation. *Curr. Biol.* 26, 1775–1781. doi: 10.1016/j.cub.2016.05.013
- Krishnaswamy, S., Verma, S., Rahman, M. H., and Kav, N. N. V. (2011). Functional characterization of four APETALA2-family genes (*RAP2.6*, *RAP2.6L*, *DREB19* and *DREB26*) in *Arabidopsis*. *Plant Mol. Biol.* 75, 107–127. doi: 10.1007/s11103-010-9711-7
- Kubota, A., Kita, S., Ishizaki, K., Nishihama, R., Yamato, K. T., and Kohchi, T. (2014). Co-option of a photoperiodic growth-phase transition system during land plant evolution. *Nat. Commun.* 5:3668. doi: 10.1038/ncomms4668
- Lavy, M., Prigge, M. J., Tao, S., Shain, S., Kuo, A., Kirchsteiger, K., et al. (2016). Constitutive auxin response in *Physcomitrella* reveals complex interactions between Aux/IAA and ARF proteins. *eLife* 5:e13325. doi: 10.7554/eLife.13325
- Lee, E., Liu, X., Eglit, Y., and Sack, F. (2013). FOUR LIPs and MYB88 conditionally restrict the G1/S transition during stomatal formation. *J. Exp. Bot.* 64, 5207–5219. doi: 10.1093/jxb/ert313
- Lex, A., Gehlenborg, N., Strobel, H., Vuilleumot, R., and Pfister, H. (2014). UpSet: visualization of intersecting sets. *IEEE Trans. Vis. Comput. Graph.* 20, 1983–1992. doi: 10.1109/TVCG.2014.2346248
- Lin, H., and Goodenough, U. W. (2007). Gametogenesis in the *Chlamydomonas reinhardtii* minus mating type is controlled by two genes, MID and MTD1. *Genetics* 176, 913–925. doi: 10.1534/genetics.106.066167
- Lin, P.-C., Lu, C.-W., Shen, B.-N., Lee, G.-Z., Bowman, J. L., Arteaga-Vazquez, M. A., et al. (2016). Identification of miRNAs and their targets in the liverwort *Marchantia polymorpha* by integrating RNA-SEQ and degradome analyses. *Plant Cell Physiol.* 57, 339–358. doi: 10.1093/pcp/pcw020
- Linde, A., Eklund, D. M., Kubota, A., Pederson, E. R. A., Holm, K., Gyllenstrand, N., et al. (2017). Early evolution of the land plant circadian clock. *New Phytol.* 216, 576–590. doi: 10.1111/nph.14487
- Liu, H., Yu, X., Li, K., Klejnot, J., Yang, H., Lisiero, D., et al. (2008). Photoexcited CRY2 interacts with CIB1 to regulate transcription and floral initiation in *Arabidopsis*. *Science* 322, 1535–1539. doi: 10.1126/science.1163927
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8
- Martín-Trillo, M., and Cubas, P. (2010). TCP genes: a family snapshot ten years later. *Trends Plant Sci.* 15, 31–39. doi: 10.1016/j.tplants.2009.11.003
- Mccarty, D. R., Hattori, T., Carson, C. B., Vasil, V., Lazar, M., and Vasil, I. K. (1991). The viviparous-1 developmental gene of maize encodes a novel transcriptional activator. *Cell* 66, 895–905. doi: 10.1016/0092-8674(91)90436-3
- Monte, I., Ishida, S., Zamarreño, A. M., Hamberg, M., Franco-Zorrilla, J. M., García-Casado, G., et al. (2018). Ligand-receptor co-evolution shaped the jasmonate pathway in land plants. *Nat. Chem. Biol.* 14, 480–488. doi: 10.1038/s41589-018-0033-4
- Mutte, S. K., Kato, H., Rothfels, C., Melkonian, M., Wong, G. K.-S., and Weijers, D. (2018). Origin and evolution of the nuclear auxin response system. *eLife* 7:e33399. doi: 10.7554/eLife.33399
- Netotea, S., Sundell, D., Street, N. R., and Hvidsten, T. R. (2014). ComPlex: conservation and divergence of co-expression networks in *A. thaliana*, *Populus* and *O. sativa*. *BMC Genom.* 15:106. doi: 10.1186/1471-2164-15-106
- Ohashi-Ito, K., and Bergmann, D. C. (2006). *Arabidopsis* FAMA Controls the Final Proliferation/Differentiation Switch during Stomatal Development. *Plant Cell* 18, 2493–2505. doi: 10.1105/tpc.106.046136
- Ohtaka, K., Hori, K., Kanno, Y., Seo, M., and Ohta, H. (2017). Primitive auxin response without TIR1 and Aux/IAA in the charophyte alga *klebsormidium nitens*. *Plant Physiol.* 174, 1621–1632. doi: 10.1104/pp.17.00274
- Oliveros, J. C. (2007). Venny. An interactive Tool for Comparing Lists with Venn Diagrams. Available: <http://bioinfo.gp.cnb.csic.es/tools/venny/index.html>

- Omelyanchuk, N. A., Wiebe, D. S., Novikova, D. D., Levitsky, V. G., Klimova, N., Gorelova, V., et al. (2017). Auxin regulates functional gene groups in a fold-change-specific manner in *Arabidopsis thaliana* roots. *Sci. Rep.* 7:2489. doi: 10.1038/s41598-017-02476-8
- Ortiz-Ramírez, C., Hernández-Coronado, M., Thamm, A., Catarino, B., Wang, M., Dolan, L., et al. (2016). A transcriptome atlas of *Physcomitrella patens* provides insights into the evolution and development of land plants. *Mol. Plant* 9, 205–220. doi: 10.1016/j.molp.2015.12.002
- Pandey, G. K. (2005). ABR1, an APETALA2-domain transcription factor that functions as a repressor of ABA response in *Arabidopsis*. *Plant Physiol.* 139, 1185–1193. doi: 10.1104/pp.105.066324
- Pedersen, D. S., Coppens, F., Ma, L., Antosch, M., Marktl, B., Merkle, T., et al. (2011). The plant-specific family of DNA-binding proteins containing three HMG-box domains interacts with mitotic and meiotic chromosomes. *New Phytol.* 192, 577–589. doi: 10.1111/j.1469-8137.2011.03828.x
- Perroud, P. F., Haas, F. B., Hiss, M., Ullrich, K. K., Alboresi, A., Amirebrahimi, M., et al. (2018). The *Physcomitrella patens* gene atlas project: large-scale RNA-seq based expression data. *Plant J.* 95, 168–192. doi: 10.1111/tpj.13940
- Proust, H. H., Honkanen, S., Jones, V. A. S. S., Morieri, G., Prescott, H., Kelly, S., et al. (2016). RSL class I genes controlled the development of epidermal structures in the common ancestor of land plants. *Curr. Biol.* 26, 93–99. doi: 10.1016/j.cub.2015.11.042
- Romani, F. (2017). Origin of TAA genes in charophytes: new insights into the controversy over the origin of auxin biosynthesis. *Front. Plant Sci.* 8:1616. doi: 10.3389/fpls.2017.01616
- Rövekamp, M., Bowman, J. L., and Grossniklaus, U. (2016). *Marchantia* MpRKD regulates the gametophyte-sporophyte transition by keeping egg cells quiescent in the absence of fertilization. *Curr. Biol.* 26, 1782–1789. doi: 10.1016/j.cub.2016.05.028
- Ruprecht, C., Mutwil, M., Saxe, F., Eder, M., Nikoloski, Z., and Persson, S. (2011). Large-scale co-expression approach to dissect secondary cell wall formation across plant species. *Front. Plant Sci.* 2:23. doi: 10.3389/fpls.2011.00023
- Ruprecht, C., and Persson, S. (2012). Co-expression of cell-wall related genes: new tools and insights. *Front. Plant Sci.* 3:83. doi: 10.3389/fpls.2012.00083
- Ruprecht, C., Proost, S., Hernandez-Coronado, M., Ortiz-Ramirez, C., Lang, D., Rensing, S. A., et al. (2017a). Phylogenomic analysis of gene co-expression networks reveals the evolution of functional modules. *Plant J.* 90, 447–465. doi: 10.1111/tpj.13502
- Ruprecht, C., Vaid, N., Proost, S., Persson, S., and Mutwil, M. (2017b). Beyond genomics: studying evolution with gene coexpression networks. *Trends Plant Sci.* 22, 298–307. doi: 10.1016/j.tplants.2016.12.011
- Schlereth, A., Moller, B., Liu, W., Kientz, M., Flipse, J., Rademacher, E. H., et al. (2010). MONOPTEROS controls embryonic root initiation by regulating a mobile transcription factor. *Nature* 464, 913–916. doi: 10.1038/nature08836
- Shimamura, M. (2016). *Marchantia polymorpha*: taxonomy, phylogeny and morphology of a model system. *Plant Cell Physiol.* 57, 230–256. doi: 10.1093/pcp/pcv192
- Singh, V. K., Rajkumar, M. S., Garg, R., and Jain, M. (2017). Genome-wide identification and co-expression network analysis provide insights into the roles of auxin response factor gene family in chickpea. *Sci. Rep.* 7:10895. doi: 10.1038/s41598-017-11327-5
- Solly, J. E., Cunniffe, N. J., and Harrison, C. J. (2017). Regional growth rate differences specified by apical notch activities regulate liverwort thallus shape. *Curr. Biol.* 27, 16–26. doi: 10.1016/j.cub.2016.10.056
- Sozzani, R., Maggio, C., Giordo, R., Umana, E., Ascencio-Ibañez, J. T., Hanley-Bowdoin, L., et al. (2010). The E2FD/DEL2 factor is a component of a regulatory network controlling cell proliferation and development in *Arabidopsis*. *Plant Mol. Biol.* 72, 381–395. doi: 10.1007/s11103-009-9577-8
- Stuart, J. M., Segal, E., Koller, D., and Kim, S. K. (2003). A gene-coexpression network for global discovery of conserved genetic modules. *Science* 302, 249–255. doi: 10.1126/science.1087447
- Sugano, S. S., Shirakawa, M., Takagi, J., Matsuda, Y., Shimada, T., Hara-Nishimura, I., et al. (2014). CRISPR/Cas9-mediated targeted mutagenesis in the liverwort *Marchantia polymorpha* L. *Plant Cell Physiol.* 55, 475–481. doi: 10.1093/pcp/pcu014
- Suzuki, M., Kao, C. Y., Cocciolone, S., and McCarty, D. R. (2001). Maize VP1 complements *Arabidopsis* abi3 and confers a novel ABA/auxin interaction in roots. *Plant J.* 28, 409–418. doi: 10.1046/j.1365-3113X.2001.01165.x
- Tsuzuki, M., Nishihama, R., Ishizaki, K., Kurihara, Y., Matsui, M., Bowman, J. L., et al. (2015). Profiling and characterization of small RNAs in the liverwort, *Marchantia polymorpha*, belonging to the first diverged land plants. *Plant Cell Physiol.* 57, 359–372. doi: 10.1093/pcp/pcv182
- Turco, G. M., Kajala, K., Kunde-Ramamoorthy, G., Ngan, C., Olson, A., Deshpande, S., et al. (2017). DNA methylation and gene expression regulation associated with vascularization in *Sorghum bicolor*. *New Phytol.* 214, 1213–1229. doi: 10.1111/nph.14448
- Wheeler, T. J., and Eddy, S. R. (2013). nhmmer: DNA homology search with profile HMMs. *Bioinformatics* 29, 2487–2489. doi: 10.1093/bioinformatics/btt403
- Yamaoka, S., Nishihama, R., Yoshitake, Y., Ishida, S., Inoue, K., Saito, M., et al. (2018). Generative cell specification requires transcription factors evolutionarily conserved in land plants. *Curr. Biol.* 28, 479–486. doi: 10.1016/j.cub.2017.12.053
- Yokoyama, A., Yamashino, T., Amano, Y. I., Tajima, Y., Imamura, A., Sakakibara, H., et al. (2007). Type-B ARR transcription factors, ARR10 and ARR12, are implicated in cytokinin-mediated regulation of protoxylem differentiation in roots of *Arabidopsis thaliana*. *Plant Cell Physiol.* 48, 84–96. doi: 10.1093/pcp/pcl040
- Yoo, S. Y., Kim, Y., Kim, S. Y., Lee, J. S., and Ahn, J. H. (2007). Control of flowering time and cold response by a NAC-domain protein in *Arabidopsis*. *PLoS One* 2:e642. doi: 10.1371/journal.pone.0000642
- Yoshii, M., Yamamoto, A., Kagaya, Y., Takeda, S., and Hattori, T. (2015). The *Arabidopsis* transcription factor NAIL1 is required for enhancing the active histone mark but not for removing the repressive mark on PYK10, a seedling-specific gene upon embryonic-to-postgerminative developmental phase transition. *Plant Signal. Behav.* 10:e1105418. doi: 10.1080/15592324.2015.1105418

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 © 2018 Flores-Sandoval, Romani and Bowman. 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 the Symbiosis-Specific GRAS Regulatory Network in Bryophytes

Christopher Grosche¹, Anne Christina Genau¹ and Stefan A. Rensing^{1,2*}

¹ Plant Cell Biology, Faculty of Biology, University of Marburg, Marburg, Germany, ² BIOS Centre for Biological Signalling Studies, University of Freiburg, Freiburg, Germany

OPEN ACCESS

Edited by:

Annette Becker,
Justus-Liebig-Universität Gießen,
Germany

Reviewed by:

Didier Reinhardt,
Université de Fribourg, Switzerland
Annegret Kohler,
INRA Centre Nancy-Lorraine, France

*Correspondence:

Stefan A. Rensing
stefan.rensing@biologie.uni-
marburg.de

Specialty section:

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

Received: 14 May 2018

Accepted: 18 October 2018

Published: 06 November 2018

Citation:

Grosche C, Genau AC and
Rensing SA (2018) Evolution of the
Symbiosis-Specific GRAS Regulatory
Network in Bryophytes.
Front. Plant Sci. 9:1621.
doi: 10.3389/fpls.2018.01621

Arbuscular mycorrhiza is one of the most common plant symbiotic interactions observed today. Due to their nearly ubiquitous occurrence and their beneficial impact on both partners it was suggested that this mutualistic interaction was crucial for plants to colonize the terrestrial habitat approximately 500 Ma ago. On the plant side the association is established via the common symbiotic pathway (CSP). This pathway allows the recognition of the fungal symbiotic partner, subsequent signaling to the nucleus, and initiation of the symbiotic program with respect to specific gene expression and cellular re-organization. The downstream part of the CSP is a regulatory network that coordinates the transcription of genes necessary to establish the symbiosis, comprising multiple GRAS transcription factors (TFs). These regulate their own expression as an intricate transcriptional network. Deduced from non-host genome data the loss of genes encoding CSP components coincides with the loss of the interaction itself. Here, we analyzed bryophyte species with special emphasis on the moss *Physcomitrella patens*, supposed to be a non-host, for the composition of the GRAS regulatory network components. We show lineage specific losses and expansions of several of these factors in bryophytes, potentially coinciding with the proposed host/non-host status of the lineages. We evaluate losses and expansions and infer clade-specific evolution of GRAS TFs.

Keywords: bryophyte, land plant evolution, moss, mycorrhiza, *Physcomitrella patens*, symbiotic pathway, GRAS, transcription factor

INTRODUCTION

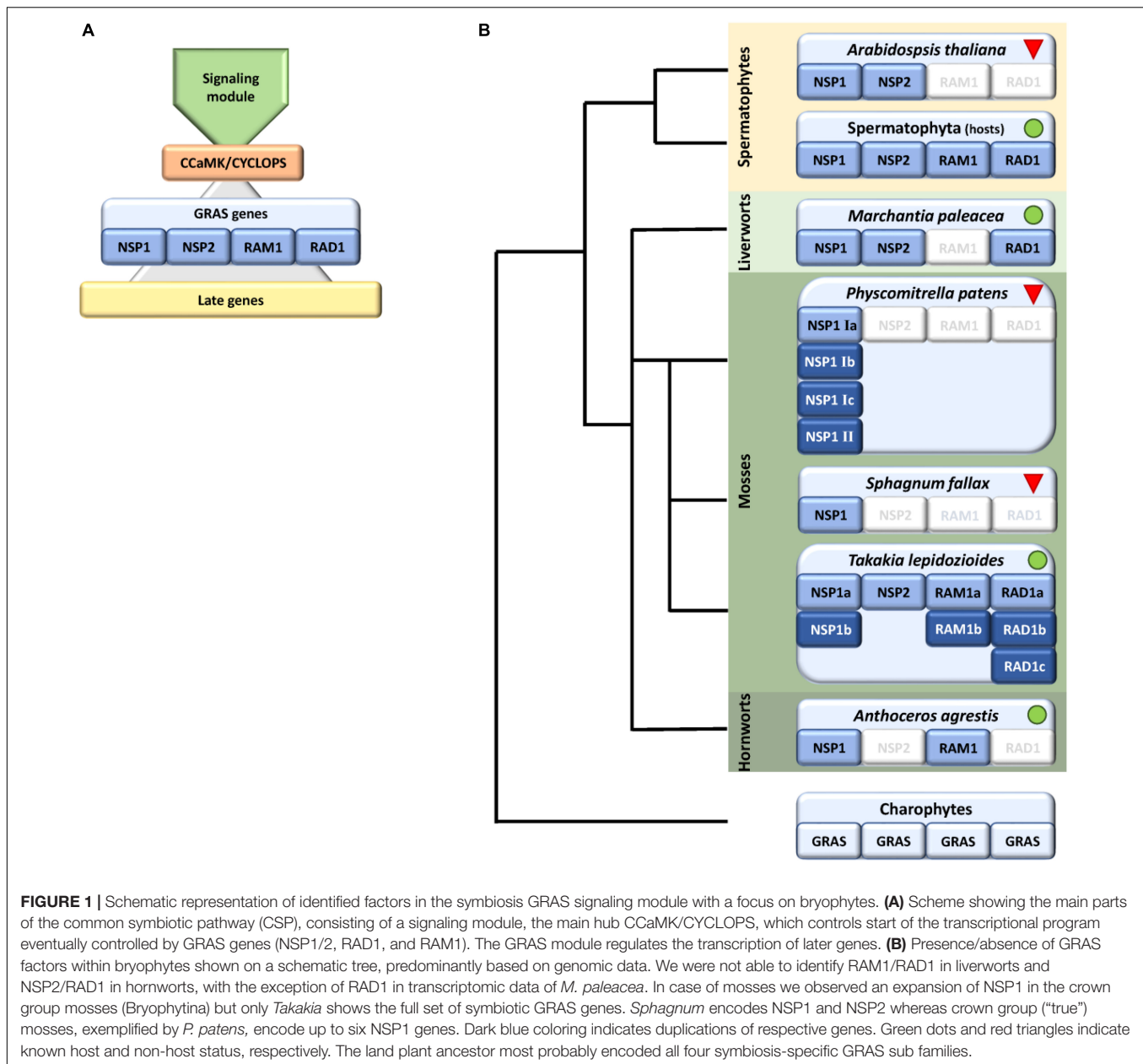
Mycorrhiza is the most common plant–fungus symbiotic interaction we observe today. Over 80% of all extant plant species engage in this symbiotic interaction (Bonfante and Genre, 2010) which is beneficial to both partners (mutualistic). The plant provides the fungus with carbohydrates and lipids, in turn the fungus provides the plant host with nutrients like nitrate and especially phosphorus. Additionally, the fungal hyphae enlarge the rhizosphere area of the plant and seem to improve plant stress tolerance (Bago et al., 2003; Liu et al., 2007; Feddermann et al., 2010; Veresoglou et al., 2012). Several forms of mycorrhizal interactions exist, of which arbuscular mycorrhiza (AM) is the most common one. It is called ‘arbuscular’ since the fungal hyphae grow

into the plant cells forming a 'tree-like' structure called arbuscule. In AM this structure represents the nutrient exchange zone between plant and fungus, since the plant-derived so-called periarbuscular membrane is loaded with transporters to facilitate the described nutrient exchange. The actual composition of the respective membrane in terms of transporters is predominantly known for plants (Harrison et al., 2002; Bonfante and Genre, 2010; Gaude et al., 2012; Luginbuehl and Oldroyd, 2017; MacLean et al., 2017).

Although beneficial for them, plants need to regulate and coordinate this symbiotic interaction because intensive cellular reprogramming is required and the plant needs to restrict the degree of colonization by the fungus in correspondence to its own nutritional status (Koide and Schreiner, 1992; Breuillin et al., 2010; Balzergue et al., 2011), e.g., to avoid carbon loss (Carbonnel and Gutjahr, 2014). Additionally, and perhaps most important, the beneficial partner needs to be distinguished from potential pathogens. Plants, and most probably already their progenitors, the streptophyte algae, evolved the so called common symbiotic pathway (CSP) (Oldroyd, 2013; Delaux et al., 2015) that enables this distinctive signaling. The pathway is called 'common' because a large part of the set of genes that evolved to accommodate arbuscular mycorrhiza (AM) was later recruited by the *Rhizobium* legume symbiosis (Kistner and Parniske, 2002). Numerous components of the CSP in AM host plants have been identified, but these analyses were predominantly performed in seed plants (see Delaux et al., 2013b; Oldroyd, 2013 for review and **Figure 1A** for overview). Via this pathway the plant detects the nearby fungus by its secreted lipo-chito-oligosaccharides (LCOs) and other myc factors such as short-chain chitin oligomers (COs) (Maillet et al., 2012; Genre et al., 2013) and prepares for colonization by starting a specific transcriptional program and cellular reorganization (Gutjahr and Parniske, 2013; Pimprikar and Gutjahr, 2018). In turn, the fungus senses the plant host, predominantly through strigolactones secreted by the plant, and starts intensive hyphal growth and branching toward the symbiotic partner (Akiyama et al., 2005; Besserer et al., 2006). The fungal signals are recognized by a receptor complex at the plant plasma membrane that involves Lysine motif (LysM) receptor like kinases (RLKs). This complex seems to be more intricate than previously thought, since it becomes more and more evident that multiple signals and receptors contribute to composite signal processing (Antolín-Llovera et al., 2012; Conn and Nelson, 2015; Gutjahr et al., 2015; Sun et al., 2015; Zhang et al., 2015). The signal is transduced to the nucleus by a so far not fully characterized mechanism involving most probably mevalonate and potentially further factors (Venkateshwaran et al., 2015). Multiple ion channels in the nuclear envelope elicit a symbiotic Ca^{2+} oscillating signal (spiking) in the nucleus (Charpentier et al., 2008). The factors described so far make up what we will henceforth call the 'signaling module' of the CSP (**Figure 1A**). This module transduces the external signal to the nucleus where it results in calcium oscillation. This symbiosis-specific calcium spiking activates the calcium and calmodulin-dependent kinase (CCaMK), a key player of the CSP, which in turn regulates the transcription factor (TF) CYCLOPS, which is thought to initiate a transcriptional regulatory network (Singh et al., 2014;

Pimprikar et al., 2016). This network of various TFs controls, together with CYCLOPS, the transcription of some additional TFs and the 'later genes' that encode factors which are, for example, needed for arbuscule initiation, branching and transmembrane transport (Harrison et al., 2002; Zhang et al., 2010; Takeda et al., 2011). The transcription of all those factors is tightly regulated and especially GRAS [Gibberellic acid insensitive (GAI), Repressor of GAI (RGA), and Scarecrow (SCR)] proteins are important regulators in this developmental process (Gutjahr, 2014; Xue et al., 2015). This family originated from a bacterial methylase (Zhang et al., 2012) and apparently evolved in streptophyte algae (Wilhelmsson et al., 2017), the sister lineage to land plants. GRAS proteins fulfill important regulatory roles in plant growth, response to environment and development (Peng et al., 1999; Pysh et al., 1999; Hirsch and Oldroyd, 2009; Sun et al., 2011). Recently the DNA binding capability of GRAS proteins was reported (Li et al., 2016), demonstrating that they might act as TFs. However, their mode of action as regulators is still highly debated (Hirano et al., 2017). In case of arbuscular mycorrhiza, so far, predominantly Reduced Arbuscular Mycorrhization 1 (RAM1), Required for Arbuscule Development1 (RAD1), Nodulation signaling pathway 1 (NSP1) and NSP2 were identified as prominent regulators, although NSP1 and NSP2 were previously thought to be root nodule symbiosis specific (Gobbato et al., 2012, 2013; Laressergues et al., 2012; Maillet et al., 2012; Delaux et al., 2013a; Hohnjec et al., 2015; Park et al., 2015; Rich et al., 2015; Xue et al., 2015; Pimprikar et al., 2016) (**Figure 1A**). Additionally, recently further potential GRAS TFs were proposed to be involved in mycorrhizal regulation (Xue et al., 2015; Heck et al., 2016). It has been suggested that the action of the four mentioned GRAS TFs is highly interconnected or dependent of each other, thus forming a transcriptional network (Xue et al., 2015). For instance, it was shown that RAM1 interacts with RAD1 and controls several 'later genes' (Park et al., 2015; Xue et al., 2015). The transcription of RAM1 in turn is controlled by CYCLOPS and DELLA (Pimprikar et al., 2016). NSP1 and NSP2 were shown to interact directly in nodulation (Hirsch et al., 2009). Additionally, NSP2 was shown to interact with RAM1 by yeast-2-hybrid and bimolecular fluorescence complementation (Gobbato et al., 2012). Adding an additional layer of complexity in the control of this symbiosis, it was shown that NSP2 is regulated by the microRNA MiR171h in flowering plants (Devers et al., 2011; Laressergues et al., 2012; Hofferek et al., 2014).

Due to their structure and mode of action, e.g., being functional as homo- or hetero-dimer, GRAS TFs seem not only to act as TFs but also as some kind of 'hub proteins' to interconnect signals from different pathways, e.g., hormone signaling, to regulate complex cellular reprogramming (Thieulin-Pardo et al., 2015; Li et al., 2016). This hub function becomes obvious in the example of DELLA proteins. These proteins are specialized GRAS TFs consisting of a GRAS domain and an additional DELLA domain. These are known to be key regulators in gibberellic acid (GA) signaling (Sun, 2011). GA presence inhibits arbuscule formation, and DELLA proteins are degraded under this condition. In turn, DELLA proteins, although not AM specific, are important for arbuscule formation



(Floss et al., 2013). Very recently it was shown that DELLAs interact with CCaMK/CYCLOPS and potentially additional TFs to regulate RAM1 transcription (Pimprikar et al., 2016). Additionally, DELLAs interact with the GRAS TF DELLA interacting protein 1 (DIP1) and RAD1, which in turn interact with RAM1 (Yu et al., 2014; Park et al., 2015; Takeda et al., 2015; Xue et al., 2015; Pimprikar et al., 2016), indicating the importance of DELLA and plant hormones in AM development and regulation. Moreover, abscisic acid (ABA) has been shown to promote mycorrhizal development, possibly by stabilizing DELLA (Achard et al., 2006), and by regulating GA levels in the context of symbiosis (Martin-Rodriguez et al., 2016).

Recently, the evolution of the CSP was analyzed, covering datasets ranging from chlorophytes to spermatophytes

(Delaux et al., 2014, 2015). It was shown that CSP factors are present in charophyte algae, especially those for signal perception and processing of the Ca^{2+} signal in the nucleus. Hence, some CSP factors were already present before the water-to-land transition (Delaux et al., 2015). However, with respect to the GRAS genes, orthologs of the symbiosis-specific GRAS TFs known from extant land plants were not detected (Delaux et al., 2015).

Symbiosis specific genes are lost in non-host plants, leading to a specific absence/presence pattern of CSP components (Delaux et al., 2014; Favre et al., 2014; Bravo et al., 2016). Hence, presence and absence of these factors may allow a conclusion on the symbiotic status of the plant analyzed. Indeed, some land plant lineages lost the ability to form a mycorrhizal partnership,

among them the Brassicaceae with the prime plant model, the weed *Arabidopsis thaliana*. The model moss *Physcomitrella patens* (Funariaceae) is not known to form AMF associations in nature, although intracellular growth can be occasionally detected in culture (Hanke and Rensing, 2010) and the relative *Funaria hygrometrica* was described to show AMF association in a companion plant assay (Parke and Linderman, 1980). While *A. thaliana* has lost genes required for responding to symbiotic fungi (Delaux et al., 2014), *P. patens* retained orthologs of these, at least for factors of the signaling module of the CSP (Wang et al., 2010; Delaux et al., 2015). Our study focuses on bryophytes, comprising mosses, hornworts, and liverworts. While liverworts and hornworts are generally considered host plants, most mosses are considered non-hosts (Field and Pressel, 2018). The crown group mosses (Bryophytina or true mosses) comprise the classes Oedipodiopsida, Polytrichopsida, Tetraphidopsida (each with a single sub class) as well as the major class Bryopsida, comprising eight sub classes. Sister lineages to the Bryophytina are the three single class comprising sub divisions Andreaephytina, Sphagnophytina (comprising the genus *Sphagnum*, peat mosses) and Takakiophytina. Their branching order remains under debate, with Takakiophytina (comprising the single genus *Takakia* with the two species *T. lepidozoides* and *T. ceratophylla*) probably being sister to all other mosses (Volkmar and Knoop, 2010; Ligrone et al., 2012). The only accepted evidence for host plants within the mosses is in the basal lineage represented by *Takakia* (Boullard, 1988). Here, we performed comprehensive phylogenomic analyses of the GRAS transcriptional regulatory network in bryophytes and found lineage specific losses and expansions of these key symbiotic signaling pathway components. We evaluate our findings with respect to the host and non-host status of the species or lineages in question and hypothesize on the (early) clade-specific evolution of symbiotic GRAS genes.

MATERIALS AND METHODS

Phylogenetic Analyses

GRAS TFs were acquired using the HMM-based (using the motif PF03514) TAPscan classification (Wilhelmsson et al., 2017) against a database of sequenced plant and algal genomes and transcriptomes (**Supplementary Table S1**). An initial alignment and phylogenetic tree of all GRAS TFs was constructed, and sequences from the clades representing the sub families involved in symbiosis signaling (NSP1, NSP2, RAD1, RAM1) were selected, aligned, manually curated and used to generate HMMs specific to each of the four sub families using hmmbuild from HMMer (Finn et al., 2015) 3.1b1 (HMMs available upon request). HMMsearch was used with these HMMs against all GRAS proteins in order to determine each of the sub families. To aid this selection, HMM search scores were derived of the basalmost sequences of the phylogenetic clade in question, and of the next closest phylogenetic clade in the tree. Cutoff scores were then derived to lie between these values. Because the resulting list of sequences of RAD1 and RAM1 largely overlapped, these two clades were combined in a single phylogenetic analysis.

While all sequences of non-vascular plants were used, seed plants were represented by selected species to cover gymnosperms, basal angiosperms as well as mono- and di-cotyledonous flowering plants. Additionally, a putative *Lunularia cruciata* RAM1 sequence (Delaux et al., 2015) was added, but based on our phylogenetic analyses could not be confirmed as RAM1. Each of the three protein sets was aligned using Mafft L-INS-i (Katoh and Standley, 2013). Alignments were manually curated using Jalview (Waterhouse et al., 2009), removing identical sequences and cropping columns to only represent the GRAS domain. Sequences of non-symbiotic GRAS proteins, namely the *A. thaliana* DELLA proteins GAI and RGA, were added for outgroup rooting (see **Supplementary Figure S1** for relationships of the GRAS clades). The best suited amino acid substitution model was determined using Prottest 3 (Darriba et al., 2011) and turned out to be JTT+G+F. Bayesian inference utilizing MrBayes 3.2 (Ronquist et al., 2012) was carried out with two hot and cold chains until the average standard deviation of split frequencies was below 0.01 and no more trend was observable. 150 trees each were discarded as burn-in. Resulting trees were visualized using FigTree 1.4.0¹. The three alignments that are the basis for the phylogenetic trees are provided as **Supplementary Files**.

Transcriptome completeness was assessed by determining the percentage of eukaryotic single copy orthologs represented as full length transcripts (**Supplementary Table S1**), as implemented in BUSCO (Simao et al., 2015). The 1KP transcriptome datasets are based on whole plants.

RESULTS

Many factors of the CSP have been identified in recent years and it has been shown that plants which do not engage in a mutualistic symbiosis with AMF lost CSP genes (Delaux et al., 2014; Favre et al., 2014; Bravo et al., 2016). Furthermore, it was shown that basic factors of the CSP, such as DMI3/CCaMK and DMI1/POLLUX, were already present in the streptophyte algae, the sister lineage of land plants (Wickett et al., 2014; Delaux et al., 2015). Most probably GRAS proteins and other TFs act in a complex way by forming homo- and heterodimers (or multimers) among each other which then regulate the respective target genes (Gobbato et al., 2012, 2013; Hohnjec et al., 2015; Xue et al., 2015). Since mycorrhiza establishment requires fundamental changes to cell structure and physiology, the transcriptional program needs to be tightly regulated and factors involved are key regulators of this alteration. This regulatory network was so far predominantly studied in spermatophytes (especially in the legumes lotus and medicago) leading to a somewhat biased knowledge about presence and absence of these factors. Therefore, we screened for homologs of GRAS TFs (NSP1, NSP2, RAM1, RAD1) previously described in the prime model organisms for mycorrhiza research, *Medicago truncatula* and/or *Lotus japonicus*, and further analyzed them by phylogenetic inference with special emphasis on bryophytes. Through this we were able to elucidate probable presence/absence patterns

¹ <http://tree.bio.ed.ac.uk/software/figtree>

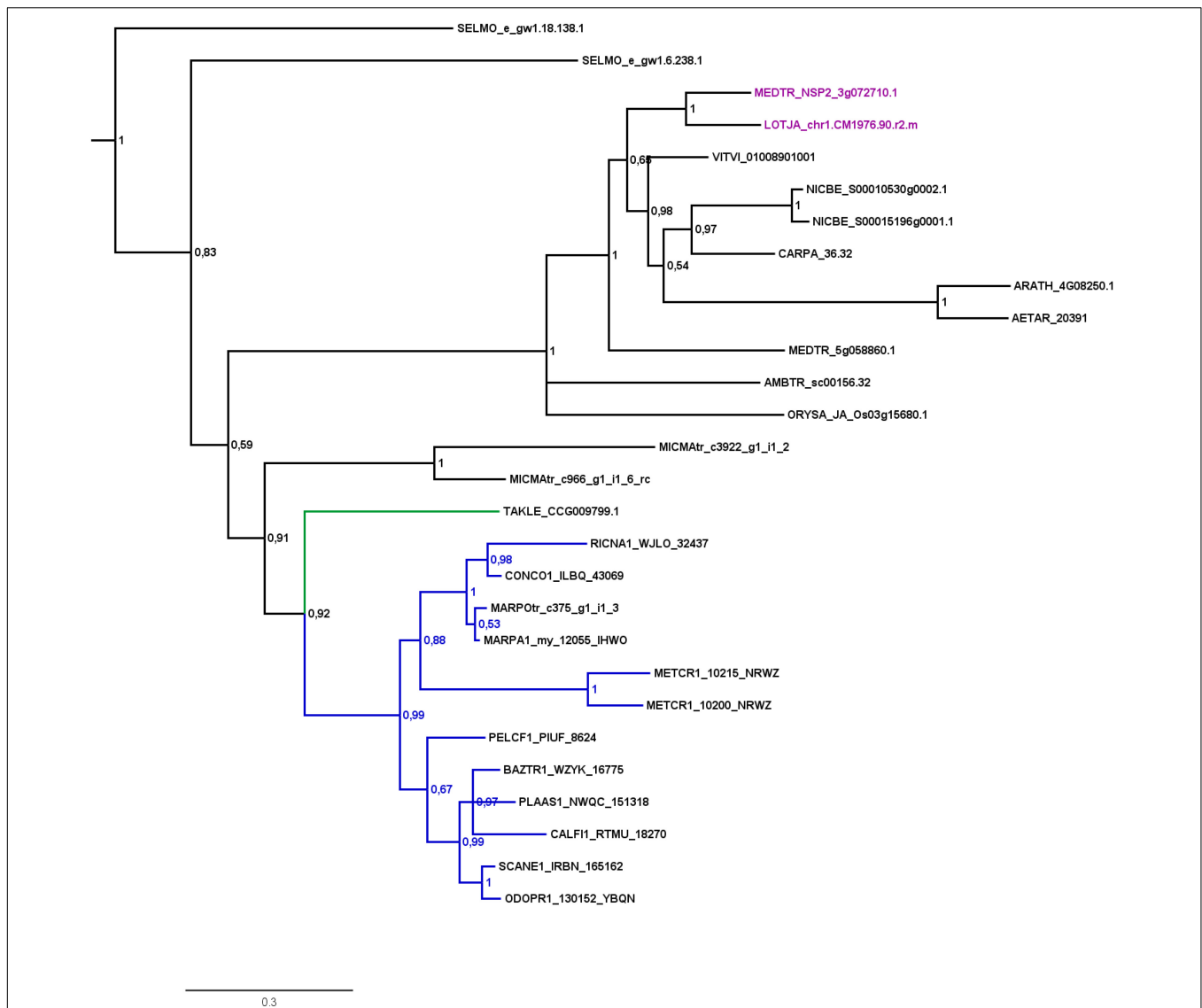


FIGURE 2 | Phylogenetic tree for NSP2. Phylogenetic reconstruction of NSP2 using Bayesian inference. NSP2 was found in the basal lineage represented by the moss *Takakia* (green branches) and in liverworts (blue branches), but not in hornworts. Sequences of *L. japonicum* and *M. truncatula* are highlighted in purple. Five letter codes of the form *MAR*chantia *PO*lymorpha (*MARPO*) are used to abbreviate species names (**Supplementary Table S1**). Five letter codes followed by “1” depict 1KP sequences, those followed by “tr” represent sequences from non-1KP transcriptomes. The *Marchantia paleacea* transcriptome generated from mycorrhized tissue is marked by “my,” the one generated from non-mycorrhized tissue by “nm.” Note that an NSP2 transcript was found in the mycorrhized (my) 1KP library for *Marchantia paleacea*, as well as in the *Marchantia polymorpha* transcriptome. Posterior probabilities are shown at the nodes, the tree was outgroup-rooted by *A. thaliana* GAI and RGA (not shown).

for factors of the GRAS transcriptional network in bryophyte clades.

Presence and Absence of Factors of the GRAS Regulatory Network in Bryophyte Lineages

Analysis of factors downstream of the signaling module and Ca^{2+} spiking (CCaMK) in non-seed plants and algae has in part already been undertaken recently (Delaux et al., 2015). Analyzing specifically the GRAS TFs and creating GRAS TF trees in more

depth (**Figures 2–4** and **Supplementary Figure S1**), we found no orthologs of NSP2 (**Figure 2**), RAD1 and RAM1 (**Figure 3**) in *P. patens* or other crown group mosses (Bryophytina) based on fully sequenced genomes, transcriptome data (Szovenyi et al., 2010, 2014) and data from the 1,000 plant transcriptomes project 1KP (Matasci et al., 2014). We also included genomic data for the more basal lineages of *Sphagnum* and *Takakia*. In case of *Sphagnum* we were able to detect only NSP1 (**Figures 1B, 4**). In contrast to that we found all factors to be present in *Takakia*.

Our GRAS phylogenetic analyzes did not detect liverwort orthologs of RAM1, although 1KP data (Matasci et al., 2014)

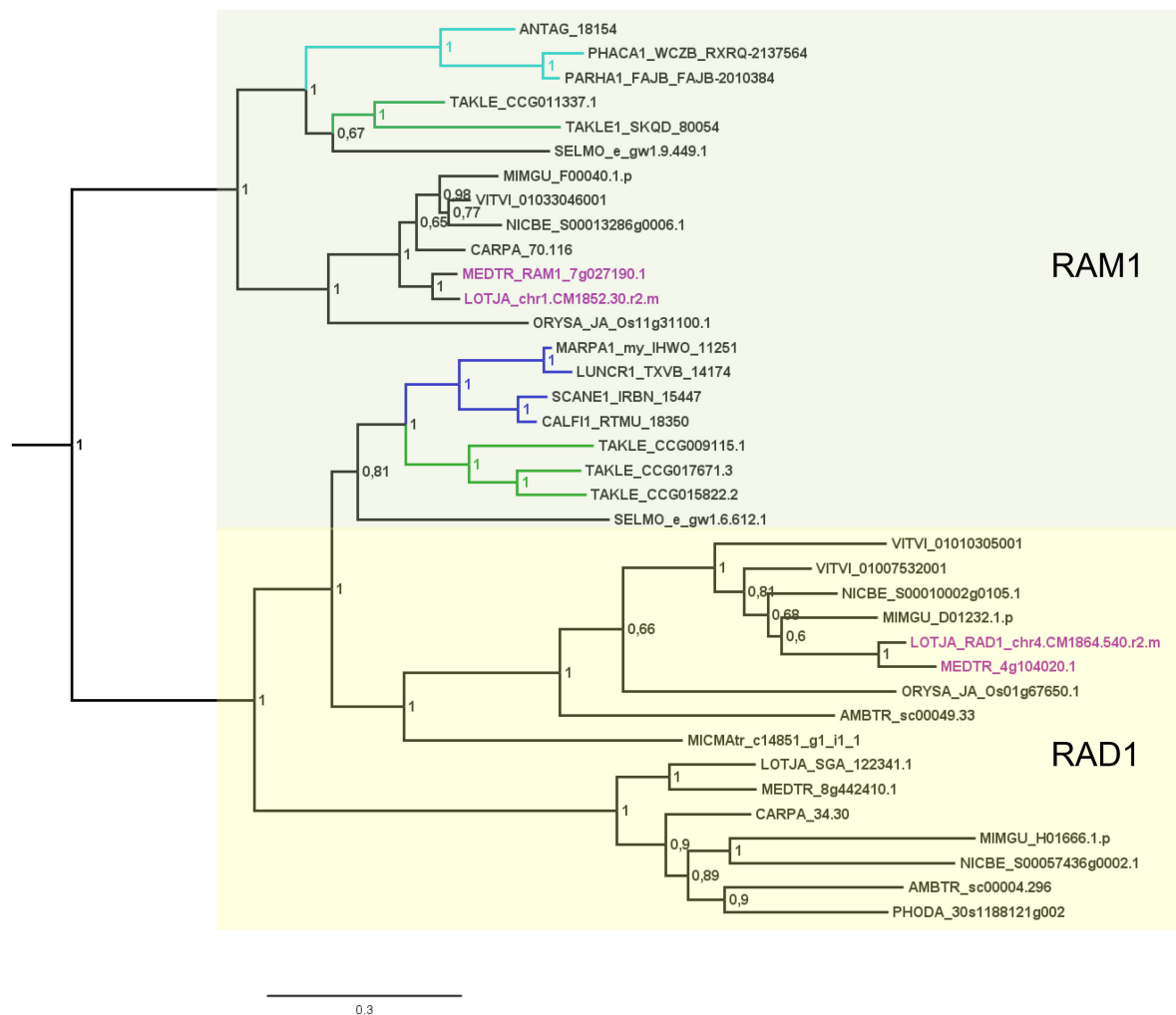


FIGURE 3 | Phylogenetic tree for RAD1 and RAM1. Phylogenetic reconstruction of RAD1 and RAM1 (clades marked using boxes) using Bayesian inference. RAD1 was found in the basal moss lineage represented by *Takakia* (green branches) and in liverworts (blue branches), but not in hornworts. In contrast, RAM1 was found only in *Takakia* and hornworts (cyan branches). Sequences of *L. japonicum* and *M. truncatula* are highlighted in purple. Note that a RAD1 transcript was only found in the mycorrhized (my) 1KP library for *Marchantia paleacea*, and that no RAD1 gene was found in the *Marchantia polymorpha* genome or transcriptome. Posterior probabilities are shown at the nodes, the tree was outgroup-rooted by *A. thaliana* GAI and RGA (not shown); see legend to **Figure 2** for explanation of naming.

were included (**Figure 3**). Apart from that we found NSP1, NSP2, and RAD1 in the liverwort lineage. Interestingly, while NSP1 and NSP2 are detected in mycorrhizal as well as non-mycorrhizal tissue of *Marchantia paleacea* and *Marchantia polymorpha*, RAD1 is only detected in mycorrhizal *M. paleacea* (**Figure 3**). For hornworts we had access to 1KP data and preliminary sequence data for *Anthoceros agrestis* (kindly provided by Peter Szovenyi). We could identify NSP1 and RAM1, but not NSP2 or RAD1 in hornworts (**Figures 1–4**). In summary, the only full set of symbiotic GRAS TFs in bryophytes was detected in genomic data of the basal moss lineage represented by *Takakia lepidiozoides* (kindly provided by Yikun He).

Lineage-Specific Expansion of GRAS TFs

Besides the mentioned lineage-specific absence of genes, our analyzes show expansions of some GRAS sub families.

Although no orthologs for NSP2, RAM1, and RAD1 were found in *P. patens*, we found four paralogs for NSP1 and a general expansion of this GRAS TF in crown group mosses (Bryophytina). Two main clades of NSP1s in mosses are obvious and most mosses seem to possess four NSP1 paralogs divided into the two clades (**Figure 4**). In case of *Sphagnum* and *Andreaea* we detected only one copy of NSP1 each, and two copies in case of *Takakia*. These three species represent the sister lineages to the Bryophytina and obviously did not share the later evolutionary diversification of NSP1. Interestingly, expansions of the other three GRAS sub families were only observable for *Takakia*, for which we identified one NSP2 but two paralogs of RAM1 and three paralogs of RAD1 (**Figure 3**).

Taken together, we identified a specific absence/presence pattern for the analyzed GRAS TFs involved in mycorrhiza signaling. Losses of parts of the symbiotic GRAS genes were

found in all examined bryophyte lineages except for *Takakia*. *P. patens* as example for the mosses shows most losses, having lost NSP2, RAM1 and RAD1. Expansions of factors were detected in *Takakia* (NSP1, RAM1, RAD1) and Bryophytina (NSP1) (**Figure 1B**). Although data of algae were included in the database mined, and GRAS TFs are present in streptophyte algae (Bowman et al., 2017; Wilhelmsson et al., 2017), orthologs of the symbiotic GRAS TFs could not be identified in algae.

DISCUSSION

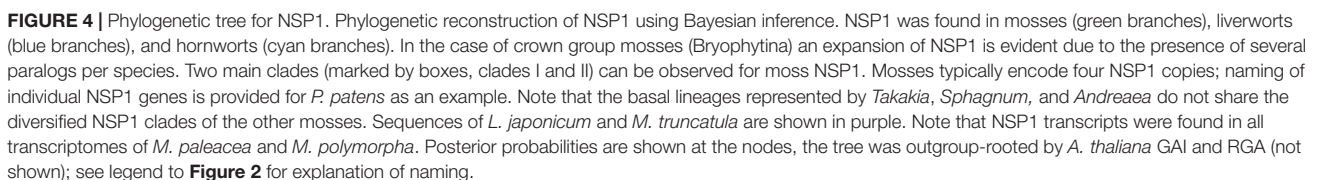
From the beginning of their conquest of land, around 500 Ma ago (Lang et al., 2010; Morris et al., 2018), land plants most probably have been in contact and/or symbiosis with fungal partners and nowadays over 80% of all extant land plants continue this mutualistic relationship (Fitter, 2005). Plants make use of a signaling module for recognition and establishment of the symbiosis, the main factors of which have already been present before the water-to-land transition of plants (Delaux et al., 2015). This might indicate that these factors are also important for microbial interactions in an aquatic environment, which are also common although so far less studied (Hempel et al., 2008; Rodriguez et al., 2009; Wurzbacher et al., 2010; Kataržytė et al., 2017). This view is supported by the fact that root nodule symbioses also make use of the CSP to establish plant–bacterial symbiosis (Oldroyd, 2013). Additionally, some microbes adopt this pathway to have parasitic access to plants; most probably parasitism is as ancient as symbiosis or predated and led to it (Corradi and Bonfante, 2012; Wang et al., 2012; Gobbato et al., 2013; Rey et al., 2015, 2017). As outlined above, components of the signaling module but not symbiosis related GRAS TFs were already present in the most recent common ancestor of land plants and charophyte algae. Indeed, we were also not able to identify sequences of charophytes orthologous to ‘symbiotic’ GRAS TFs. As proposed before (Delaux et al., 2015) the symbiotic GRAS signaling most probably evolved by duplication events from GRAS TFs already present in streptophyte algae (Bowman et al., 2017; Wilhelmsson et al., 2017).

Symbiotic GRAS TFs in Bryophytes – Duplications and Losses

Delaux et al. (2015) detected orthologs for symbiotic GRAS TFs in bryophytes. The full complement was detected for liverworts only (especially *Lunularia cruciata*, for which the transcriptome was sequenced). Furthermore, they identified two factors (RAM1 and RAD1) for *Takakia*, and NSP1 in hornworts and mosses. Our GRAS phylogenetic trees (**Figures 2–4**) expand this view in so far that we found all four factors in *Takakia*, added the basal moss lineage represented by *Sphagnum* (having NSP1 only), and found NSP2 and RAM1 in hornworts. We were not able to identify RAM1 in liverworts. The previously reported putative *Lunularia cruciata* transcriptomic RAM1 sequence (Delaux et al., 2015) grouped outside the RAM1/RAD1 clade in our analysis, maybe due to its fragmentary sequence. The detection of these GRAS TFs in transcriptomic data is potentially flawed because some of them are only expressed upon detection of or colonization

by AM fungi, and thereby activation of the CSP (Xue et al., 2015; Pimprikar et al., 2016; Rich et al., 2017). Hence, they might not be expressed under the conditions from which the respective transcriptome was sequenced. Additionally, hornworts are unfortunately underrepresented in the 1KP data. Such problems do not apply if full genomes (with a certain quality) are available (e.g., *P. patens* or the liverwort *M. polymorpha*). The strength of our study is that we use for the first time genomic data for each of the bryophyte lineages, thus at least partially overcoming the limitations of transcriptomic data. However, it also clearly demonstrates that we need more genomic data for bryophytes and other non-seed plants (Rensing, 2017).

In the case of mosses, only NSP1 is present and clearly expanded, exemplified by, e.g., *P. patens* encoding four NSP1 genes. Using the *M. truncatula* sequences encoding NSP2, RAD1 and RAM1, no hits can be recovered in the *P. patens* genome assembly. Also, the best BLASTP hits of the *M. truncatula* genes flanking the three GRAS loci are not part of syntenic regions detected between *P. patens* and other plant genomes (Lang et al., 2018). Neither is any of the four NSP1 paralogs part of such a syntenic region. Hence, the regions encoding three of the four genes seem to have been lost from the genome. As outlined above, there are two moss NSP1 subclades (**Figure 4**, clades I and II). Most of the mosses have at least one sequence in each of these clades, and typically encode four paralogs. The topology and distribution pattern of NSP1 genes indicate a common duplication event giving rise to the two clades, observed in the crown group mosses (Bryophytina) but not shared by the sister lineages represented by *Takakia*, *Sphagnum*, and *Andreaea*. These duplications might be related to whole genome duplication (WGD) events observable in mosses (Lang et al., 2018), leading to subsequent neo- and/or sub-functionalization of duplicated genes (Rensing, 2014). Published expression for *P. patens* (Perroud et al., 2018) shows that the four paralogs show different expression levels (Nsp1 Ib lowest, Nsp1 Ia highest), but a qualitatively similar expression profile across the available developmental stages (**Supplementary Table S1**). NSP1 was identified as being important in root nodule symbiosis (RNS) (Catoira et al., 2000; Smit et al., 2005) and later on it was shown to also influence AM, since this TF is an important factor of the strigolactone (SL) biosynthesis pathway (Liu et al., 2011; Delaux et al., 2013b; Takeda et al., 2013; Hohnjec et al., 2015). SL biosynthesis is important for the establishment of the AM symbiosis because the fungus senses SL and the hyphal branching increases upon this stimulus (Akiyama et al., 2010). Interestingly, although *P. patens* does not encode an NSP2 gene, which is also necessary for SL biosynthesis in seed plants (Liu et al., 2011), it releases SLs (Proust et al., 2011). Biosynthesis of SL in *P. patens* is induced by phosphate starvation (a condition under which AMF association typically occurs), and leads to resistance to pathogenic fungi (Decker et al., 2017). Potentially, duplicated moss NSP1 genes sub-/neo-functionalized and compensate for the loss of NSP2. As GRAS proteins can act as hetero- and/or homodimers (Hirsch et al., 2009; Li et al., 2016) it is a feasible scenario that the four NSP1 paralogs might take over functions typically carried out by other GRAS proteins in other plants. Intriguingly, only NSP1 genes of clade I (**Figure 4**) seem to have been duplicated



in mosses (the naming of individual NSP1 genes in the overview in **Figure 1B** is according to this division). This might indicate that a first sub- or neo-functionalization of NSP1 genes already occurred after the first duplication and in the second duplication event duplicated genes of group II were selected against, maybe because of unfavorable consequences due to stoichiometry of the dimer partners. However, so far, we are not able to assign certain functions to the individual NSP1 genes in *P. patens*.

Presence/Absence of Factors Coinciding With the Host/Non-host Status

Looking at the potential overall evolution of the symbiosis GRAS signaling genes we found support for the view of Delaux et al. (2015). Given the distribution of GRAS factors we hypothesize that the land plant ancestor encoded the full set of GRAS TFs (**Figure 1**). In bryophyte clades we can observe several losses (especially in mosses) and expansions (also mainly in mosses) of these GRAS TFs. Mosses have lost most factors of these GRAS genes (NSP2, RAM1, RAD1) and according to that they, including *Sphagnum*, are considered non-host plants (Read et al., 2000; Wang and Qiu, 2006). Although the signaling module seems to be intact, they also lost some genes which are, e.g., important for the periarbuscular membrane (Wang et al., 2010; Delaux et al., 2014, 2015). The symbiotic GRAS sub family losses might explain why a tight association or even symbiosis cannot be established in mosses. An exception at the basis of the mosses is *Takakia*, which encodes all four GRAS factors (plus expansions for NSP1, RAM1, and RAD1), and indeed was reported to engage in AM (Boullard, 1988). Given that *Takakia* represents one of the sister lineages of the Bryophytina, an evolutionary loss of NSP2, RAD1, and RAM1 during moss evolution appears the most probable scenario.

Most liver- and hornworts are considered host plants (Field and Pressel, 2018). As mentioned, we were not able to identify RAM1 in liverworts, or NSP2 and RAD1 in hornworts. This might be due to the mentioned problem with transcriptome coverage, but is also explainable by species-specific losses of AM capability exemplified by *M. polymorpha*, which does not show AM and lacks RAD1 (**Figure 3**) in contrast to its close relative *M. paleacea* (Humphreys et al., 2010; Bowman et al., 2017). The lack of *M. polymorpha* mycorrhizal association is in line with the absence of GAs and might be featured by nutrient rich habitats (Ligrone et al., 2007). While genes needed for successful mycorrhization are absent in non-host *Marchantia* species, other gene families are over-represented in *M. polymorpha*, e.g., transporters for phosphate and ammonium. These genomic adaptations might reflect the shift from mycorrhizal to non-mycorrhizal status by improving the transport capacity instead of being dependent on symbiotic organisms (Bowman et al., 2017). Nevertheless, according to our analysis both lack RAM1, which is not in line with the species' host/non-host status, since *M. paleacea*, as host, should have the complete GRAS gene set. This is most probably due to the mentioned incomplete nature of the transcriptomic data. In case of NSP2 we can identify a potential coding region in the *M. polymorpha* genome

(encoding the same protein detected in the transcriptome, **Figure 2**) that does not have a gene model assigned to it; updated genome versions might solve this issue. Presence of transcripts in transcriptomic data is evidence of presence of the gene, but absence of transcripts must not necessarily reflect absence of the gene (for example, the 1KP transcriptomes contain on average 84% of the conserved eukaryotic single copy gene set, **Supplementary Table S1**). For our overview (**Figure 1B**) we are thus relying mainly on genomic data in order not to represent conclusions based on transcriptomic absence of genes.

Functions, Additional Factors and Evolution of the Symbiotic Pathway

With rising morphological complexity more complex regulation and cellular reorganization is needed. Most probably we do not yet know all TFs involved in the regulation of this symbiosis, which involves tight regulation and massive cellular reorganization. Recent publications indicate that even more factors are involved, at least in seed plants (Xue et al., 2015; Heck et al., 2016). This indicates that we are only at the beginning of understanding this complex pathway and its transcriptional network (Genre and Russo, 2016; Pimprikar and Gutjahr, 2018). However, a quick survey showed that for example the GRAS TF MIG1 seems not to be present in bryophytes (data not shown). Most probably the transcriptional network to establish mycorrhizal symbiosis comprises more factors and is thereby more complex in vascular plants, due to more cell types and tissue layers. It is important to note that our current knowledge of the CSP is predominantly based on studies in spermatophytes (and here again predominantly analyzed of lotus and medicago). Maybe some GRAS genes are less important in bryophytes as compared to the situation in seed or vascular plants, for example due to their lack of roots.

If we evaluate the evolution of the CSP and its downstream components we should also be considering other plant-microbe associations and symbioses. Recently, the view was broadened since fungi belonging to the Mucoromycotina and also Ascomycotina were shown to interact in particular with liverworts and maybe hornworts (Field et al., 2014; Kowal et al., 2018), and the plant-Mucoromycotina interaction was proposed to potentially represent the ancestral state of plant-fungus interaction (Field et al., 2015). It is also known that plants interact with bacteria or even with both, fungi and bacteria (Bonfante and Anca, 2009). Foremost known is the RNS, which makes use of many factors of the CSP (Oldroyd, 2013; Genre and Russo, 2016). Interaction with cyanobacteria is also known, in particular for hornworts but also in some liverworts, mosses, ferns, and seed plants. There are more microbial/plant associations and symbioses known (e.g., *Frankia*, etc.) (Santi et al., 2013; Martin et al., 2017), and most probably many more we do not know yet. These interactions are important and widespread, and probably evolved already in the aquatic environment (Croft et al., 2005; Hom and Murray, 2014). How are these associations and symbioses regulated and how do the symbiotic partners identify each other? Most likely key

components of the CSP (signaling module) are also involved in symbioses other than the ones they have so far been implicated in (mycorrhizal, rhizobial, and actinorhizal) (Martin et al., 2017). The signaling module apparently evolved in streptophyte algae (Delaux et al., 2015), suggesting that it may be functional in additional associations and symbioses, e.g., with cyanobacteria. The factors that process the symbiotic calcium spiking and induce a specific transcriptional program might be specific for each kind of symbiosis, leading to an association/symbiosis-specific diversification of the CSP downstream of the signaling module. More molecular studies in additional symbioses, especially in aquatic environments, are needed to unravel additional symbiosis-specific factors.

CONCLUSION

The key pathway to regulate beneficial interactions in plants seems to be the CSP (Martin et al., 2017). Furthermore, it is believed that microbial interactions enabled plants to conquer the land (Pirozynski and Malloch, 1975), highlighting the importance of this signaling pathway. Here we argue that the symbiosis related GRAS signaling genes, known to be important in regulation of AM symbiosis, were already present in the most recent common ancestor of all land plants. These genes display lineage specific losses and expansions in bryophytes, in particular in mosses. Such losses seem to reflect the non-host status. Nevertheless, the upstream CSP signaling module for symbiosis establishment seems to be intact (Wang et al., 2010; Delaux et al., 2015) and may serve in additional symbioses with a different or an extended subset of factors in the transcriptional network module. Additional studies are needed to elucidate the symbiosis-specific interplay of TFs and the functions of, e.g., the duplicated NSP1 genes in mosses.

AUTHOR CONTRIBUTIONS

SAR and CG conceived of the study. SAR carried out the phylogenetic analyses. CG, ACG, and SAR analyzed the

data. CG and SAR wrote the paper, with contributions by ACG.

FUNDING

This project was funded by the German Research Foundation (DFG RE1697/6-1 to SAR), the Forschungsförderfond of the University of Marburg, the University of Freiburg/the Ministry of Science, Research and Art of the state of Baden-Württemberg (RiSC co-grant to SAR).

ACKNOWLEDGMENTS

The authors thank M. Göttig for excellent technical assistance, C. Gutjahr for critical comments on the manuscript, and N. Fernandez-Pozo for carrying out BUSCO analyses. They would like to thank Peter Szovenyi for access to *Anthoceros agrestis* and Yikun He for access to *Takakia lepidiozoides* draft genome data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.01621/full#supplementary-material>

FIGURE S1 | Overview phylogeny of the GRAS family. Midpoint rooted Bayesian inference tree of selected species, including *Lotus japonicus*, *Medicago truncatula*, *Arabidopsis thaliana*, *Physcomitrella patens*, and *Marchantia polymorpha*. Line thickness corresponds to posterior probabilities. Colored clades depict RAD1 (red), RAM1 (purple), NSP1 (blue), NSP2 (green), and DELLA (cyan). Note that DELLA proteins were used as outgroup to root each of the trees shown in **Figures 2–4**.

TABLE S1 | Lists the sources of genomic and transcriptomic datasets used, five letter species abbreviations used in the Figures, BUSCO completeness percentages and expression data for the *P. patens* NSP1 paralogs. Alignments of NSP1, NSP2, and RAD1/RAM1 used for the phylogenetic analyses shown in **Figures 2–4**.

REFERENCES

- Achard, P., Cheng, H., De Grauwe, L., Decat, J., Schoutteten, H., Moritz, T., et al. (2006). Integration of plant responses to environmentally activated phytohormonal signals. *Science* 311, 91–94. doi: 10.1126/science.1118642
- Akiyama, K., Matsuzaki, K., and Hayashi, H. (2005). Plant sesquiterpenes induce hyphal branching in arbuscular mycorrhizal fungi. *Nature* 435, 824–827. doi: 10.1038/nature03608
- Akiyama, K., Ogasawara, S., Ito, S., and Hayashi, H. (2010). Structural requirements of strigolactones for hyphal branching in AM fungi. *Plant Cell Physiol.* 51, 1104–1117. doi: 10.1093/pcp/pcq058
- Antolín-Llovera, M., Ried, M. K., Binder, A., and Parniske, M. (2012). Receptor kinase signaling pathways in plant-microbe interactions. *Annu. Rev. Phytopathol.* 50, 451–473. doi: 10.1146/annurev-phyto-081211-173002
- Bago, B., Pfeffer, P. E., Abubaker, J., Jun, J., Allen, J. W., Brouillette, J., et al. (2003). Carbon export from arbuscular mycorrhizal roots involves the translocation of carbohydrate as well as lipid. *Plant Physiol.* 131, 1496–1507. doi: 10.1104/pp.102.007765
- Balzergrue, C., Puech-Pages, V., Bécard, G., and Rochange, S. F. (2011). The regulation of arbuscular mycorrhizal symbiosis by phosphate in pea involves early and systemic signalling events. *J. Exp. Bot.* 62, 1049–1060. doi: 10.1093/jxb/erq335
- Besserer, A., Puech-Pages, V., Kiefer, P., Gomez-Roldan, V., Jauneau, A., Roy, S., et al. (2006). Strigolactones stimulate arbuscular mycorrhizal fungi by activating mitochondria. *PLoS Biol.* 4:e226. doi: 10.1371/journal.pbio.0040226
- Bonfante, P., and Anca, I. A. (2009). Plants, mycorrhizal fungi, and bacteria: a network of interactions. *Annu. Rev. Microbiol.* 63, 363–383. doi: 10.1146/annurev.micro.091208.073504
- Bonfante, P., and Genre, A. (2010). Mechanisms underlying beneficial plant–fungus interactions in mycorrhizal symbiosis. *Nat. Commun.* 1:48. doi: 10.1038/ncomms1046
- Boullard, B. (1988). “Observations on the coevolution of fungi with hepatics,” in *Coevolution of Fungi with Plants and Animals*, eds K. A. Pirozynski and D. L. Hawksworth (Cambridge, MA: Academic Press).

- 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.e15. doi: 10.1016/j.cell.2017.09.030
- Bravo, A., York, T., Pumplun, N., Mueller, L. A., and Harrison, M. J. (2016). Genes conserved for arbuscular mycorrhizal symbiosis identified through phylogenomics. *Nat. Plants* 2:15208. doi: 10.1038/nplants.2015.208
- Breullin, F., Schramm, J., Hajirezaei, M., Ahkami, A., Favre, P., Druge, U., et al. (2010). Phosphate systemically inhibits development of arbuscular mycorrhiza in *Petunia hybrida* and represses genes involved in mycorrhizal functioning. *Plant J.* 64, 1002–1017. doi: 10.1111/j.1365-313X.2010.04385.x
- Carbonnel, S., and Gutjahr, C. (2014). Control of arbuscular mycorrhiza development by nutrient signals. *Front. Plant Sci.* 5:462. doi: 10.3389/fpls.2014.00462
- Catoira, R., Galera, C., de Billy, F., Penmetsa, R. V., Journet, E. P., Maillet, F., et al. (2000). Four genes of *Medicago truncatula* controlling components of a nod factor transduction pathway. *Plant Cell* 12, 1647–1666. doi: 10.1105/tpc.12.9.1647
- Charpentier, M., Bredemeier, R., Wanner, G., Takeda, N., Schleiff, E., and Parniske, M. (2008). *Lotus japonicus* CASTOR and POLLUX are ion channels essential for perinuclear calcium spiking in legume root endosymbiosis. *Plant Cell* 20, 3467–3479. doi: 10.1105/tpc.108.063255
- Conn, C. E., and Nelson, D. C. (2015). Evidence that KARRIKIN-INSENSITIVE2 (KAI2) receptors may perceive an unknown signal that is not karrikin or strigolactone. *Front. Plant Sci.* 6:1219. doi: 10.3389/fpls.2015.01219
- Corradi, N., and Bonfante, P. (2012). The arbuscular mycorrhizal symbiosis: origin and evolution of a beneficial plant infection. *PLoS Pathog.* 8:e1002600. doi: 10.1371/journal.ppat.1002600
- Croft, M. T., Lawrence, A. D., Raux-Deery, E., Warren, M. J., and Smith, A. G. (2005). Algae acquire vitamin B12 through a symbiotic relationship with bacteria. *Nature* 438, 90–93. doi: 10.1038/nature04056
- Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27, 1164–1165. doi: 10.1093/bioinformatics/btr088
- Decker, E. L., Alder, A., Hunn, S., Ferguson, J., Lehtonen, M. T., Scheler, B., et al. (2017). Strigolactone biosynthesis is evolutionarily conserved, regulated by phosphate starvation and contributes to resistance against phytopathogenic fungi in a moss, *Physcomitrella patens*. *New Phytol.* 216, 455–468. doi: 10.1111/nph.14506
- Delaux, P.-M., Bécard, G., and Combier, J.-P. (2013a). NSP1 is a component of the Myc signaling pathway. *New Phytol.* 199, 59–65. doi: 10.1111/nph.12340
- Delaux, P.-M., Séjalon-Delmas, N., Bécard, G., and Ané, J.-M. (2013b). Evolution of the plant-microbe symbiotic 'toolkit'. *Trends Plant Sci.* 18, 298–304. doi: 10.1016/j.tplants.2013.01.008
- Delaux, P. M., Radhakrishnan, G. V., Jayaraman, D., Cheema, J., Malbreil, M., Volkening, J. D., et al. (2015). Algal ancestor of land plants was preadapted for symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* 112, 13390–13395. doi: 10.1073/pnas.1515426112
- Delaux, P. M., Varala, K., Edger, P. P., Coruzzi, G. M., Pires, J. C., and Ane, J. M. (2014). Comparative phylogenomics uncovers the impact of symbiotic associations on host genome evolution. *PLoS Genet.* 10:e1004487. doi: 10.1371/journal.pgen.1004487
- Devers, E. A., Branscheid, A., May, P., and Krajinski, F. (2011). Stars and symbiosis: microRNA- and microRNA*-mediated transcript cleavage involved in arbuscular mycorrhizal symbiosis. *Plant Physiol.* 156, 1990–2010. doi: 10.1104/pp.111.172627
- Favre, P., Bapaume, L., Bossolini, E., Delorenzi, M., Falquet, L., and Reinhardt, D. (2014). A novel bioinformatics pipeline to discover genes related to arbuscular mycorrhizal symbiosis based on their evolutionary conservation pattern among higher plants. *BMC Plant Biol.* 14:333. doi: 10.1186/s12870-014-0333-0
- Feddermann, N., Finlay, R., Boller, T., and Elfstrand, M. (2010). Functional diversity in arbuscular mycorrhiza – the role of gene expression, phosphorous nutrition and symbiotic efficiency. *Fungal Ecol.* 3, 1–8. doi: 10.1016/j.funeco.2009.07.003
- Field, K. J., and Pressel, S. (2018). Unity in diversity: structural and functional insights into the ancient partnerships between plants and fungi. *New Phytol.* doi: 10.1111/nph.15158 [Epub ahead of print].
- Field, K. J., Pressel, S., Duckett, J. G., Rimington, W. R., and Bidartondo, M. I. (2015). Symbiotic options for the conquest of land. *Trends Ecol. Evol.* 30, 477–486. doi: 10.1016/j.tree.2015.05.007
- Field, K. J., Rimington, W. R., Bidartondo, M. I., Allinson, K. E., Beerling, D. J., Cameron, D. D., et al. (2014). First evidence of mutualism between ancient plant lineages (Haplomitriopsida liverworts) and Mucoromycotina fungi and its response to simulated Palaeozoic changes in atmospheric CO₂. *New Phytol.* 205, 743–756. doi: 10.1111/nph.13024
- Finn, R. D., Clements, J., Arndt, W., Miller, B. L., Wheeler, T. J., Schreiber, F., et al. (2015). HMMER web server: 2015 update. *Nucleic Acids Res.* 43, W30–W38. doi: 10.1093/nar/gkv397
- Fitter, A. H. (2005). Darkness visible: reflections on underground ecology. *J. Ecol.* 93, 231–243. doi: 10.1111/j.0022-0477.2005.00990.x
- Floss, D. S., Levy, J. G., Levesque-Tremblay, V., Pumplun, N., and Harrison, M. J. (2013). DELLA proteins regulate arbuscule formation in arbuscular mycorrhizal symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* 110, E5025–E5034. doi: 10.1073/pnas.1308973110
- Gaude, N., Schulze, W. X., Franken, P., and Krajinski, F. (2012). Cell type-specific protein and transcription profiles implicate periarbuscular membrane synthesis as an important carbon sink in the mycorrhizal symbiosis. *Plant Signal. Behav.* 7, 461–464. doi: 10.4161/psb.19650
- Genre, A., Chabaud, M., Balzergue, C., Puech-Pages, V., Novero, M., Rey, T., et al. (2013). Short-chain chitin oligomers from arbuscular mycorrhizal fungi trigger nuclear Ca²⁺ spiking in *Medicago truncatula* roots and their production is enhanced by strigolactone. *New Phytol.* 198, 190–202. doi: 10.1111/nph.12146
- Genre, A., and Russo, G. (2016). Does a common pathway transduce symbiotic signals in plant-microbe interactions? *Front. Plant Sci.* 7:96. doi: 10.3389/fpls.2016.00096
- Gobbato, E., Marsh, J. F., Vernié, T., Wang, E., Maillet, F., Kim, J., et al. (2012). A GRAS-type transcription factor with a specific function in mycorrhizal signaling. *Curr. Biol.* 22, 2236–2241. doi: 10.1016/j.cub.2012.09.044
- Gobbato, E., Wang, E., Higgins, G., Bano, S. A., Henry, C., Schultze, M., et al. (2013). RAM1 and RAM2 function and expression during arbuscular mycorrhizal symbiosis and *Aphanomyces euteiches* colonization. *Plant Signal. Behav.* 8:e26049. doi: 10.4161/psb.26049
- Gutjahr, C. (2014). Phytohormone signaling in arbuscular mycorrhiza development. *Curr. Opin. Plant Biol.* 20, 26–34. doi: 10.1016/j.pbi.2014.04.003
- Gutjahr, C., Gobbato, E., Choi, J., Riemann, M., Johnston, M. G., Summers, W., et al. (2015). Rice perception of symbiotic arbuscular mycorrhizal fungi requires the karrikin receptor complex. *Science* 350, 1521–1524. doi: 10.1126/science.aac9715
- Gutjahr, C., and Parniske, M. (2013). Cell and developmental biology of arbuscular mycorrhiza symbiosis. *Annu. Rev. Cell Dev. Biol.* 29, 593–617. doi: 10.1146/annurev-cellbio-101512-122413
- Hanke, S. T. and Rensing, S. A. (2010). *In vitro* association of non-seed plant gametophytes with arbuscular mycorrhiza fungi. *Endocyt. Cell. Res.* 20, 95–101.
- Harrison, M. J., Dewbre, G. R., and Liu, J. (2002). A phosphate transporter from *Medicago truncatula* involved in the acquisition of phosphate released by arbuscular mycorrhizal fungi. *Plant Cell* 14, 2413–2429. doi: 10.1105/tpc.004861
- Heck, C., Kuhn, H., Heidt, S., Walter, S., Rieger, N., and Requena, N. (2016). Symbiotic fungi control plant root cortex development through the novel GRAS transcription factor MIG1. *Curr. Biol.* 26, 2770–2778. doi: 10.1016/j.cub.2016.07.059
- Hempel, M., Blume, M., Blindow, I., and Gross, E. M. (2008). Epiphytic bacterial community composition on two common submerged macrophytes in brackish water and freshwater. *BMC Microbiol.* 8:58. doi: 10.1186/1471-2180-8-58
- Hirano, Y., Nakagawa, M., Suyama, T., Murase, K., Shirakawa, M., Takayama, S., et al. (2017). Structure of the SHR-SCR heterodimer bound to the BIRD/IDD transcriptional factor JKD. *Nat. Plants* 3:17010. doi: 10.1038/nplants.2017.10
- Hirsch, S., Kim, J., Munoz, A., Heckmann, A. B., Downie, J. A., and Oldroyd, G. E. (2009). GRAS proteins form a DNA binding complex to induce gene expression during nodulation signaling in *Medicago truncatula*. *Plant Cell* 21, 545–557. doi: 10.1105/tpc.108.064501
- Hirsch, S., and Oldroyd, G. E. (2009). GRAS-domain transcription factors that regulate plant development. *Plant Signal. Behav.* 4, 698–700. doi: 10.4161/psb.4.8.9176
- Hofferek, V., Mendrinna, A., Gaude, N., Krajinski, F., and Devers, E. A. (2014). MiR171h restricts root symbioses and shows like its target NSP2 a complex

- transcriptional regulation in *Medicago truncatula*. *BMC Plant Biol.* 14:199. doi: 10.1186/s12870-014-0199-1
- Hohnjec, N., Czaja-Hasse, L. F., Hoge Kamp, C., and Kuster, H. (2015). Pre-announcement of symbiotic guests: transcriptional reprogramming by mycorrhizal lipochitoooligosaccharides shows a strict co-dependency on the GRAS transcription factors NSP1 and RAM1. *BMC Genomics* 16:994. doi: 10.1186/s12864-015-2224-7
- Hom, E. F., and Murray, A. W. (2014). Plant-fungal ecology. Niche engineering demonstrates a latent capacity for fungal-algal mutualism. *Science* 345, 94–98. doi: 10.1126/science.1253320
- Humphreys, C. P., Franks, P. J., Rees, M., Bidartondo, M. I., Leake, J. R., and Beerling, D. J. (2010). Mutualistic mycorrhiza-like symbiosis in the most ancient group of land plants. *Nat. Commun.* 1:103. doi: 10.1038/ncomms1105
- Kataržytė, M., Vaičiūtė, D., Bučas, M., Gyraitė, G., and Petkuvienė, J. (2017). Microorganisms associated with charophytes under different salinity conditions. *Oceanologia* 59, 177–186. doi: 10.1016/j.oceano.2016.10.002
- 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
- Kistner, C., and Parniske, M. (2002). Evolution of signal transduction in intracellular symbiosis. *Trends Plant Sci.* 7, 511–518. doi: 10.1016/S1360-1385(02)02356-7
- Koide, R. T., and Schreiner, R. P. (1992). Regulation of the vesicular-arbuscular mycorrhizal symbiosis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 43, 557–581. doi: 10.1146/annurev.pp.43.060192.003013
- Kowal, J., Pressel, S., Duckett, J. G., Bidartondo, M. I., and Field, K. J. (2018). From rhizoids to roots? Experimental evidence of mutualism between liverworts and ascomycete fungi. *Ann. Bot.* 121, 221–227. doi: 10.1093/aob/mcx126
- Lang, D., Ullrich, K. K., Murat, F., Fuchs, J., Jenkins, J., Haas, F. B., et al. (2018). The *Physcomitrella patens* chromosome-scale assembly reveals moss genome structure and evolution. *Plant J.* 93, 515–533. doi: 10.1111/tpj.13801
- Lang, D., Weiche, B., Timmerhaus, G., Richardt, S., Riano-Pachon, D. M., Correa, L. G., et al. (2010). Genome-wide phylogenetic comparative analysis of plant transcriptional regulation: a timeline of loss, gain, expansion, and correlation with complexity. *Genome Biol. Evol.* 2, 488–503. doi: 10.1093/gbe/evq032
- Lauressergues, D., Delaux, P. M., Formey, D., Lelandais-Briere, C., Fort, S., Cottaz, S., et al. (2012). The microRNA miR171h modulates arbuscular mycorrhizal colonization of *Medicago truncatula* by targeting NSP2. *Plant J.* 72, 512–522. doi: 10.1111/j.1365-313X.2012.05099.x
- Li, S., Zhao, Y., Zhao, Z., Wu, X., Sun, L., Liu, Q., et al. (2016). Crystal structure of the GRAS domain of SCARECROW-LIKE7 in *Oryza sativa*. *Plant Cell* 28, 1025–1034. doi: 10.1105/tpc.16.00018
- Ligrone, R., Carafa, A., Lumini, E., Bianciotto, V., Bonfante, P., and Duckett, J. G. (2007). Glomeromycotene associations in liverworts: a molecular, cellular, and taxonomic analysis. *Am. J. Bot.* 94, 1756–1777. doi: 10.3732/ajb.94.11.1756
- 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
- Liu, J., Maldonado-Mendoza, I., Lopez-Meyer, M., Cheung, F., Town, C. D., and Harrison, M. J. (2007). Arbuscular mycorrhizal symbiosis is accompanied by local and systemic alterations in gene expression and an increase in disease resistance in the shoots. *Plant J.* 50, 529–544. doi: 10.1111/j.1365-313X.2007.03069.x
- Liu, W., Kohlen, W., Lillo, A., Op, den Camp, R., Ivanov, S., et al. (2011). Strigolactone biosynthesis in *Medicago truncatula* and rice requires the symbiotic GRAS-type transcription factors NSP1 and NSP2. *Plant Cell Online* 23, 3853–3865. doi: 10.1105/tpc.111.089771
- Luginbuehl, L. H., and Oldroyd, G. E. D. (2017). Understanding the arbuscule at the heart of endomycorrhizal symbioses in plants. *Curr. Biol.* 27, R952–R963. doi: 10.1016/j.cub.2017.06.042
- MacLean, A. M., Bravo, A., and Harrison, M. J. (2017). Plant signaling and metabolic pathways enabling arbuscular mycorrhizal symbiosis. *Plant Cell* 29, 2319–2335. doi: 10.1105/tpc.17.00555
- Maillet, F., Poinot, V., André, O., Puech-Pages, V., Haouy, A., Gueunier, M., et al. (2012). Fungal lipochitoooligosaccharide symbiotic signals in arbuscular mycorrhiza. *Nature* 469, 58–63. doi: 10.1038/nature09622
- Martin, F. M., Uroz, S., and Barker, D. G. (2017). Ancestral alliances: plant mutualistic symbioses with fungi and bacteria. *Science* 356:eaad4501.
- Martin-Rodriguez, J. A., Huertas, R., Ho-Plagaro, T., Ocampo, J. A., Tureckova, V., Tarkowska, D., et al. (2016). Gibberellin-abscisic acid balances during arbuscular mycorrhiza formation in tomato. *Front. Plant Sci.* 7:1273. doi: 10.3389/fpls.2016.01273
- 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* 3:17. doi: 10.1186/2047-217X-3-17
- Morris, J. L., Puttick, M. N., Clark, J. W., Edwards, D., Kenrick, P., Pressel, S., et al. (2018). The timescale of early land plant evolution. *Proc. Natl. Acad. Sci. U.S.A.* 115, E2274–E2283. doi: 10.1073/pnas.1719588115
- Oldroyd, G. E. (2013). Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nat. Rev. Microbiol.* 11, 252–263. doi: 10.1038/nrmicro2990
- Park, H.-J., Floss, D. S., Levesque-Tremblay, V., Bravo, A., and Harrison, M. J. (2015). Hyphal branching during arbuscule development requires reduced arbuscular mycorrhiza1. *Plant Physiol.* 169, 2774–2788.
- Parke, J. L., and Linderman, R. G. (1980). Association of vesicular-arbuscular mycorrhizal fungi with the moss *Funaria hygrometrica*. *Can. J. Bot.* 58, 1898–1904. doi: 10.1139/b80-218
- Peng, J., Richards, D. E., Hartley, N. M., Murphy, G. P., Devos, K. M., Flintham, J. E., et al. (1999). 'Green revolution' genes encode mutant gibberellin response modulators. *Nature* 400, 256–261. doi: 10.1038/22307
- Perroud, P. F., Haas, F. B., Hiss, M., Ullrich, K. K., Alboresi, A., Amirebrahimi, M., et al. (2018). The *Physcomitrella patens* gene atlas project: large scale RNA-seq based expression data. *Plant J.* 95, 168–182. doi: 10.1111/tpj.13940
- Pimprikar, P., Carbonnel, S., Paries, M., Katzer, K., Klingl, V., Bohmer, M. J., et al. (2016). A CCA-MK-CYCLOPS-DELLA complex activates transcription of Ram1 to regulate arbuscule branching. *Curr. Biol.* 26, 987–998. doi: 10.1016/j.cub.2016.01.069
- Pimprikar, P., and Gutjahr, C. (2018). Transcriptional regulation of arbuscular mycorrhiza development. *Plant Cell Physiol.* 59, 673–690. doi: 10.1093/pcp/pcy024
- Pirozynski, K. A., and Malloch, D. W. (1975). The origin of land plants: a matter of mycotrophism. *Biosystems* 6, 153–164. doi: 10.1016/0303-2647(75)90023-4
- Proust, H., Hoffmann, B., Xie, X., Yoneyama, K., Schaefer, D. G., Yoneyama, K., et al. (2011). Strigolactones regulate protonema branching and act as a quorum sensing-like signal in the moss *Physcomitrella patens*. *Development* 138, 1531–1539. doi: 10.1242/dev.058495
- Pysh, L. D., Wysocka-Diller, J. W., Camilleri, C., Bouchez, D., and Benfey, P. N. (1999). The GRAS gene family in Arabidopsis: sequence characterization and basic expression analysis of the SCARECROW-LIKE genes. *Plant J.* 18, 111–119. doi: 10.1046/j.1365-313X.1999.00431.x
- Read, D. J. D., Duckett, J. G. J., Francis, R. R., Ligrone, R. R., and Russell, A. A. (2000). Symbiotic fungal associations in 'lower' land plants. *Philos. Trans. R. Soc. B Biol. Sci.* 355, 815–811. doi: 10.1098/rstb.2000.0617
- Rensing, S. A. (2014). Gene duplication as a driver of plant morphogenetic evolution. *Curr. Opin. Plant Biol.* 17, 43–48. doi: 10.1016/j.pbi.2013.11.002
- Rensing, S. A. (2017). Why we need more non-seed plant models. *New Phytol.* 216, 355–360. doi: 10.1111/nph.14464
- Rey, T., Bonhomme, M., Chatterjee, A., Gavrin, A., Toulotte, J., Yang, W., et al. (2017). The *Medicago truncatula* GRAS protein RAD1 supports arbuscular mycorrhiza symbiosis and *Phytophthora palmivora* susceptibility. *J. Exp. Bot.* 68, 5871–5881. doi: 10.1093/jxb/erx398
- Rey, T., Chatterjee, A., Buttay, M., Toulotte, J., and Schornack, S. (2015). *Medicago truncatula* symbiosis mutants affected in the interaction with a biotrophic root pathogen. *New Phytol.* 206, 497–500. doi: 10.1111/nph.13233
- Rich, M. K., Courty, P. E., Roux, C., and Reinhardt, D. (2017). Role of the GRAS transcription factor ATA/RAM1 in the transcriptional reprogramming of arbuscular mycorrhiza in *Petunia hybrida*. *BMC Genomics* 18:589. doi: 10.1186/s12864-017-3988-8
- Rich, M. K., Schorderet, M., Bapaume, L., Falquet, L., Morel, P., Vandenbussche, M., et al. (2015). The petunia GRAS transcription factor ATA/RAM1 regulates symbiotic gene expression and fungal morphogenesis in arbuscular mycorrhiza. *Plant Physiol.* 168, 788–797. doi: 10.1104/pp.15.00310
- Rodriguez, R. J., White, J. F. Jr., Arnold, A. E., and Redman, R. S. (2009). Fungal endophytes: diversity and functional roles. *New Phytol.* 182, 314–330. doi: 10.1111/j.1469-8137.2009.02773.x

- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029
- Santi, C., Bogusz, D., and Franche, C. (2013). Biological nitrogen fixation in non-legume plants. *Ann. Bot.* 111, 743–767. doi: 10.1093/aob/mct048
- Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351
- Singh, S., Katzer, K., Lambert, J., Cerri, M., and Parniske, M. (2014). CYCLOPS, a DNA-binding transcriptional activator, orchestrates symbiotic root nodule development. *Cell Host Microbe* 15, 139–152. doi: 10.1016/j.chom.2014.01.011
- Smit, P., Raedts, J., Portyanko, V., Debellé, F., Gough, C., Bisseling, T., et al. (2005). NSP1 of the GRAS protein family is essential for rhizobial Nod factor-induced transcription. *Science* 308, 1789–1791. doi: 10.1126/science.1111025
- Sun, J., Miller, J. B., Granqvist, E., Wiley-Kalil, A., Gobbato, E., Maillet, F., et al. (2015). Activation of symbiosis signaling by arbuscular mycorrhizal fungi in legumes and rice. *Plant Cell* 27, 823–838. doi: 10.1105/tpc.114.131326
- Sun, T. P. (2011). The molecular mechanism and evolution of the GA-GID1-DELLA signaling module in plants. *Curr. Biol.* 21, R338–R345. doi: 10.1016/j.cub.2011.02.036
- Sun, X., Xue, B., Jones, W. T., Rikkerink, E., Dunker, A. K., and Uversky, V. N. (2011). A functionally required unfoldome from the plant kingdom: intrinsically disordered N-terminal domains of GRAS proteins are involved in molecular recognition during plant development. *Plant Mol. Biol.* 77, 205–223. doi: 10.1007/s11103-011-9803-z
- Szovenyi, P., Perroud, P. F., Symeonidi, A., Stevenson, S., Quatrano, R. S., Rensing, S. A., et al. (2014). De novo assembly and comparative analysis of the *Ceratodon purpureus* transcriptome. *Mol. Ecol. Resour.* 15, 203–215. doi: 10.1111/1755-0998.12284
- Szovenyi, P., Rensing, S. A., Lang, D., Wray, G. A., and Shaw, A. J. (2010). Generation-biased gene expression in a bryophyte model system. *Mol. Biol. Evol.* 28, 803–812. doi: 10.1093/molbev/msq254
- Takeda, N., Haage, K., Sato, S., Tabata, S., and Parniske, M. (2011). Activation of a *Lotus japonicus* subtilase gene during arbuscular mycorrhiza is dependent on the common symbiosis genes and two cis-active promoter regions. *Mol. Plant Microbe Interact.* 24, 662–670. doi: 10.1094/MPMI-09-10-0220
- Takeda, N., Handa, Y., Tsuzuki, S., Kojima, M., Sakakibara, H., and Kawaguchi, M. (2015). Gibberellins interfere with symbiosis signaling and gene expression and alter colonization by arbuscular mycorrhizal fungi in *Lotus japonicus*. *Plant Physiol.* 167, 545–557. doi: 10.1104/pp.114.247700
- Takeda, N., Tsuzuki, S., Suzuki, T., Parniske, M., and Kawaguchi, M. (2013). CERBERUS and NSP1 of *Lotus japonicus* are common symbiosis genes that modulate arbuscular mycorrhiza development. *Plant Cell Physiol.* 54, 1711–1723. doi: 10.1093/pcp/pct114
- Thieulin-Pardo, G., Avilan, L., Kojadinovic, M., and Gontero, B. (2015). Fairy “tails”: flexibility and function of intrinsically disordered extensions in the photosynthetic world. *Front. Mol. Biosci.* 2:23. doi: 10.3389/fmolb.2015.00023
- Venkateshwaran, M., Jayaraman, D., Chabaud, M., Genre, A., Balloon, A. J., Maeda, J., et al. (2015). A role for the mevalonate pathway in early plant symbiotic signaling. *Proc. Natl. Acad. Sci. U.S.A.* 112, 9781–9786. doi: 10.1073/pnas.1413762112
- Veresoglou, S. D., Chen, B., and Rillig, M. C. (2012). Arbuscular mycorrhiza and soil nitrogen cycling. *Soil Biol. Biochem.* 46, 53–62. doi: 10.1016/j.soilbio.2011.11.018
- Volkmar, U., and Knoop, V. (2010). Introducing intron locus *cox1i624* for phylogenetic analyses in Bryophytes: on the issue of *Takakia* as sister genus to all other extant mosses. *J. Mol. Evol.* 70, 506–518. doi: 10.1007/s00239-010-9348-9
- Wang, B., and Qiu, Y. L. (2006). Phylogenetic distribution and evolution of mycorrhizas in land plants. *Mycorrhiza* 16, 299–363. doi: 10.1007/s00572-005-0033-6
- Wang, B., Yeun, L. H., Xue, J.-Y., Liu, Y., Ané, J.-M., and Qiu, Y.-L. (2010). Presence of three mycorrhizal genes in the common ancestor of land plants suggests a key role of mycorrhizas in the colonization of land by plants. *New Phytol.* 186, 514–525. doi: 10.1111/j.1469-8137.2009.03137.x
- Wang, E., Schornack, S., Marsh, J. F., Gobbato, E., Schwessinger, B., Eastmond, P., et al. (2012). A common signaling process that promotes mycorrhizal and oomycete colonization of plants. *Curr. Biol.* 22, 2242–2246. doi: 10.1016/j.cub.2012.09.043
- Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191. doi: 10.1093/bioinformatics/btp033
- Wickett, N. J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., et al. (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
- Wilhelmsson, P. K. I., Muhlich, C., Ullrich, K. K., and Rensing, S. A. (2017). Comprehensive genome-wide classification reveals that many plant-specific transcription factors evolved in streptophyte algae. *Genome Biol. Evol.* 9, 3384–3397. doi: 10.1093/gbe/evx258
- Wurzbacher, C. M., Bärlocher, F., and Grossart, H. P. (2010). Fungi in lake ecosystems. *Aquat. Microb. Ecol.* 59, 125–149. doi: 10.3354/ame01385
- Xue, L., Cui, H., Buer, B., Vijayakumar, V., Delaux, P. M., Junkermann, S., et al. (2015). Network of GRAS transcription factors involved in the control of arbuscule development in *Lotus japonicus*. *Plant Physiol.* 167, 854–871. doi: 10.1104/pp.114.255430
- Yu, N., Luo, D., Zhang, X., Liu, J., Wang, W., Jin, Y., et al. (2014). A DELLA protein complex controls the arbuscular mycorrhizal symbiosis in plants. *Cell Res.* 24, 130–133. doi: 10.1038/cr.2013.167
- Zhang, D., Iyer, L. M., and Aravind, L. (2012). Bacterial GRAS domain proteins throw new light on gibberellic acid response mechanisms. *Bioinformatics* 28, 2407–2411. doi: 10.1093/bioinformatics/bts464
- Zhang, Q., Blaylock, L. A., and Harrison, M. J. (2010). Two *Medicago truncatula* half-ABC transporters are essential for arbuscule development in arbuscular mycorrhizal symbiosis. *Plant Cell* 22, 1483–1497. doi: 10.1105/tpc.110.074955
- Zhang, X., Dong, W., Sun, J., Feng, F., Deng, Y., He, Z., et al. (2015). The receptor kinase CERK1 has dual functions in symbiosis and immunity signalling. *Plant J.* 81, 258–267. doi: 10.1111/tpj.12723

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 © 2018 Grosche, Genau and Rensing. 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.



Getting to the Roots: A Developmental Genetic View of Root Anatomy and Function From Arabidopsis to Lycophytes

*Frauke Augstein and Annelie Carlsbecker**

Department of Organismal Biology, Physiological Botany and Linnean Centre for Plant Biology in Uppsala, Uppsala University, Uppsala, Sweden

OPEN ACCESS

Edited by:

Annette Becker,
Justus Liebig Universität Gießen,
Germany

Reviewed by:

Hongchang Cui,
Florida State University, United States
Keiko Sakakibara,
Rikkyo University, Japan

*Correspondence:

Annelie Carlsbecker
annelie.carlsbecker@ebc.uu.se

Specialty section:

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

Received: 30 June 2018

Accepted: 05 September 2018

Published: 25 September 2018

Citation:

Augstein F and Carlsbecker A (2018)
Getting to the Roots:
A Developmental Genetic View
of Root Anatomy and Function From
Arabidopsis to Lycophytes.
Front. Plant Sci. 9:1410.
doi: 10.3389/fpls.2018.01410

Roots attach plants to the ground and ensure efficient and selective uptake of water and nutrients. These functions are facilitated by the morphological and anatomical structures of the root, formed by the activity of the root apical meristem (RAM) and consecutive patterning and differentiation of specific tissues with distinct functions. Despite the importance of this plant organ, its evolutionary history is not clear, but fossils suggest that roots evolved at least twice, in the lycophyte (clubmosses and their allies) and in the euphyllophyte (ferns and seed plants) lineages. Both lycophyte and euphyllophyte roots grow indeterminately by the action of an apical meristem, which is protected by a root cap. They produce root hairs, and in most species the vascular stele is guarded by a specialized endodermal cell layer. Hence, most of these traits must have evolved independently in these lineages. This raises the question if the development of these apparently analogous tissues is regulated by distinct or homologous genes, independently recruited from a common ancestor of lycophytes and euphyllophytes. Currently, there are few studies of the genetic and molecular regulation of lycophyte and fern roots. Therefore, in this review, we focus on key regulatory networks that operate in root development in the model angiosperm *Arabidopsis*. We describe current knowledge of the mechanisms governing RAM maintenance as well as patterning and differentiation of tissues, such as the endodermis and the vasculature, and compare with other species. We discuss the importance of comparative analyses of anatomy and morphology of extant and extinct species, along with analyses of gene regulatory networks and, ultimately, gene function in plants holding key phylogenetic positions to test hypotheses of root evolution.

Keywords: roots, plant evo-devo, plant development, plant anatomy and morphology, patterning, gene regulatory network

FOSSILS, PHYLOGENIES, AND DEVELOPMENTAL GENETICS IN TRACING THE EVOLUTION OF ROOTS

Roots anchor plants to the ground, and their growth patterns allow exploration of the soil while their specific morphology and anatomy are adapted for efficient uptake of water and mineral nutrients. The evolution of deeply penetrating roots dramatically altered living conditions on Earth. Their activity is capable of weathering rocks resulting in accessible silicate material which

reacts with and binds carbon dioxide thereby reducing it from the atmosphere (Raven and Edwards, 2001; Pires and Dolan, 2012). Roots are essential for the formation of complex soils and allow intimate symbiotic relationships with fungi and bacteria (Raven and Edwards, 2001). Thus, both the abiotic and biotic environment were altered with the evolution of roots and plant roots remain essential for our ecosystems. Furthermore, optimal plant root behavior in response to external conditions such as mineral nutrient and water availability is essential for crop survival and yield (Ahmed et al., 2018). Despite their importance, the evolutionary history of roots is currently not clearly understood and the developmental genetic regulation of root traits is known in considerable detail essentially only in the model plant *Arabidopsis thaliana* (Arabidopsis), although knowledge from other angiosperm (primarily crop) species is rapidly catching up.

All extant vascular plants have true roots (with few exceptions) distinguished by positive gravitropism and a root cap protecting a meristem that allows continuous growth. Roots have root hairs extending the surface area for efficient water and mineral uptake, and a ground tissue that often harbors an inner specialized endodermal cell layer controlling uptake into the vascular stele. The stele generally has a primitive protostele organization with central xylem (Raven and Edwards, 2001). Although being united by these characteristics the fossil record along with certain developmental features (see below) strongly suggest that roots evolved independently several times, implying that several root specific structures convergently evolved multiple times. This raises the question if distinct genetic components were employed for similar functions or if these multiple independent events involved adoption of related genetic circuits present in the ancestor of these plants. If so, they would display deep homology, i.e., when the structures themselves are analogous but regulated by homologous genes (Scotland, 2010). Exploring the gene regulatory networks underlying root development in phylogenetically informative species both at great evolutionary distances, and in closely related species with similar or distinct anatomies, will give valuable information on the genetic tool kit(s) employed in root formation, and, potentially, in their evolution. Here, we review current hypotheses of how roots might have evolved, we describe selected key genetic circuits essential for different aspects of root development, with a focus on meristem maintenance and anatomy, and discuss how such information can help testing hypotheses for root evolution.

THE ORIGIN AND ANATOMICAL DIVERSITY OF ROOTS

Rooting structures are found in all land plants; in the form of rhizoids (uni- or multicellular filamentous rooting structures emanating from non-root organs) in the free-living gametophytes of bryophytes, lycophytes, and monilophytes, and as true roots with elaborate tissues, as described above or with some variation, in the sporophytes of extant vascular plants (Raven and Edwards, 2001). This seemingly indicates a monophyletic origin of true roots. However, the fossils found of early vascular plants and

certain aspects of how roots develop in different lineages, instead suggest a considerably more complex evolutionary history of roots (Figure 1). At the time when the first root-like structures appear in the fossil record, sporophytic plants had indeterminately growing upright or crawling, rhizome-like, shoot axes, some with microphyllous leaves (Taylor et al., 2009; Kenrick and Strullu-Derrien, 2014). The Rhyniophytes found in the Rhynie chert from Early Devonian are among the earliest fossils found of sporophytes with rooting structures (Taylor et al., 2009). These distinct fossils had a system of shoot-like axes, either a pro-vasculature or a distinguishable xylem, but no roots as specified by a root meristem covered by a cap. Instead, they had rhizoids (otherwise only known from gametophytes) directly developed from the lower surfaces of the axes at places where they were growing horizontally, on top or just beneath the soil surface (Taylor et al., 2009; Kenrick and Strullu-Derrien, 2014; Hetherington and Dolan, 2018a). Thus, it is possible that these axes had adapted a specific genetic program normally responsible for gametophyte rhizoid formation, to allow something of a rooting function to these axes. Such a combination of rhizoids formed on a sporophytic axis is an early evolutionary elaboration not found in any extant plant (Hetherington and Dolan, 2018a), but it emphasizes the similarities of rhizoids and root hairs (single cell tubular epidermal outgrowths of true roots). Indeed, comparisons of the genetic regulation of bryophyte gametophyte rhizoid and angiosperm root hair formation have identified a number of homologous factors regulating the formation of the analogous tip-growing cells with rooting function (Menand et al., 2007; Tam et al., 2015; Honkanen et al., 2017, reviewed in Jones and Dolan, 2012; Honkanen and Dolan, 2016), suggesting that a pre-existing genetic network for rhizoid formation was co-opted in the sporophyte generation.

Thus, Rhyniophytes, stem group taxa for all extant vascular plants, lacked true roots. Interestingly, also fossils of stem groups of both lycophytes (Zosterophylls) and fern and seed plants (Trimerophytes) lacked true roots (Raven and Edwards, 2001; Taylor et al., 2009; Kenrick and Strullu-Derrien, 2014). Therefore, true roots likely evolved independently in these two major vascular plant branches. The earliest fossils with true roots have affinity to the lycophytes. Most of these had apparently indeterminate growth, a root cap-like structure, no cuticula, and branched dichotomously by bifurcation of the meristem and most likely lacked an endodermis. Intriguingly, a very early fossil of a lyopsid root meristem found in the Rhynie chert was recently described in great detail (Hetherington and Dolan, 2018b). While these roots had positive gravitropism and a promeristem that had set off cells for vascular, ground and epidermal tissues, they had no signs of having had a root cap. Hence, Hetherington and Dolan take this as evidence for step-wise acquisitions of key root traits during root evolution in the lycophyte lineage. Indirectly, this further supports that such root traits must have convergently evolved independently in the other major vascular plant lineage, the euphyllphytes (Hetherington and Dolan, 2018b).

In other fossils of Early Devonian lycophytes, non-gravitropic thin root-like structures in position of leaves originated from shoot-like positively gravitropic axes (Matsunaga and Tomescu, 2016). Thus, these lycophyte roots formed as a novel type of

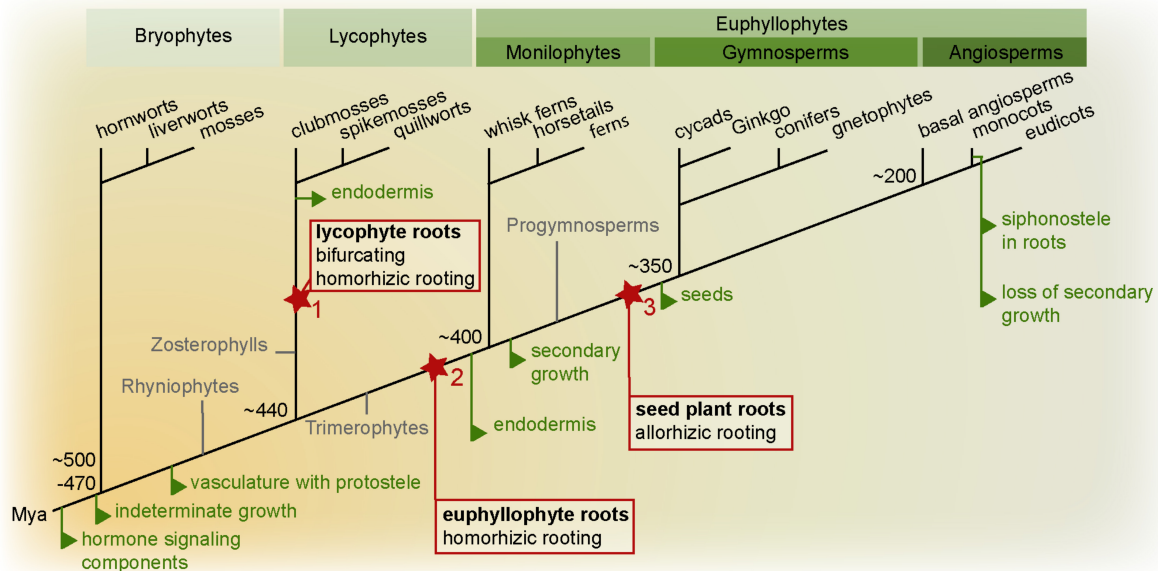


FIGURE 1 | Land plant phylogeny highlighting key root evolutionary events. Important evolutionary innovations are mapped on a phylogenetic tree of land plants (simplified tree after Puttick et al., 2018). True roots are hypothesized to have evolved at least twice, in the lycophyte (1) and euphyllophyte lineage (2). In the lycophyte lineage, the root meristem is bifurcating, while in the euphyllophyte lineage, lateral roots emerge from the endodermis or pericycle. In lycophytes and monilophytes, the root develops as an adventitious organ “homorhizic roots,” while in seed plants, the root and shoot meristems are bipolar, and the root is said to be “allorhizic” (although seed plants also form adventitious roots). It is currently unclear how homorhizic and allorhizic roots relate to each other and whether the evolution of a primary root in the seed plant lineage can be seen as a third root-evolution event (3) (see Liu and Xu, 2018). Evolution of homorhizic and allorhizic roots are marked in red, other major root evolutionary events are indicated in green. Mya: million years ago.

organ, and this fossil further suggests that certain lycophyte roots may have co-opted positive gravitropism at a later evolutionary step. In other early fossils, roots developed from many different sites of the plant: from stems, leaves, or other places (Taylor et al., 2009; Hetherington and Dolan, 2017). Hence, the diversity in where roots appeared may even suggest that roots evolved multiple times among the then very diverse lycophytes. Still today, lycophyte (*Lycopodiales*, clubmosses; *Selaginetales*, spikemosses; *Isoetales*, quillworts) roots display ancestral characters such as meristem branching by bifurcation, and *Lycopodium* has no root endodermis, although other lycophytes develop an endodermis (Hetherington and Dolan, 2017). Indeed, even among extant lycophytes there is a large variation in root meristem morphology – some have elaborate meristems with multiple stem cells as in seed plants (see below), while others have only one apical cell dividing to give rise to all root tissues as in ferns – supporting the paleobotanical indications of multiple evolution of roots in the lycophyte lineage (Fujinami et al., 2017).

In contrast to the lycophytes, euphyllophytes do not branch by bifurcation, but from internal tissues (from the endodermis in ferns and pericycle in seed plants) proximally to the apical meristem. Similar to lycophytes, fern roots develop as adventitious outgrowths in relation to the longitudinal axis of the embryo, and form so called “homorhizic” roots (Raven and Edwards, 2001). In fossils (e.g., *Archaeopteris*) of progymnosperms, from which the seed plants evolved, plants have been found where the root formed at the opposing end

of the shoot apical meristem (SAM), referred to as “bipolar” or “allorhizic” roots (Taylor et al., 2009). All seed plants are distinguished by having allorhizic roots although it is not uncommon to also find roots developing from non-root organs such as stems or leaves. It is currently an unresolved issue if the allorhizic seed plant root is homologous with the homorhizic fern root.

Thus, the evolutionary history of roots is still not clear, but there is good evidence that roots appeared as multiple independent innovations in the two major lineages leading to extant lycophytes and euphyllophytes. One can envision different trajectories by which roots may have evolved: They may have appeared as an entirely new type of organ, perhaps as a modification of a lateral shoot organ, or they could have evolved as modifications of shoot-like axes, already harboring an apical meristem.

THE ROOT APICAL MERISTEM AND AUXIN – SIMILARITIES AND DIFFERENCES IN ANGIOSPERMS, FERNS, AND LYCOPHYTES

As discussed above, the evolution of roots was predated by shoots growing indeterminate. Thus, genetic circuits ensuring indeterminate growth must have been present, and may have been co-opted to allow indeterminate growth of roots. This could

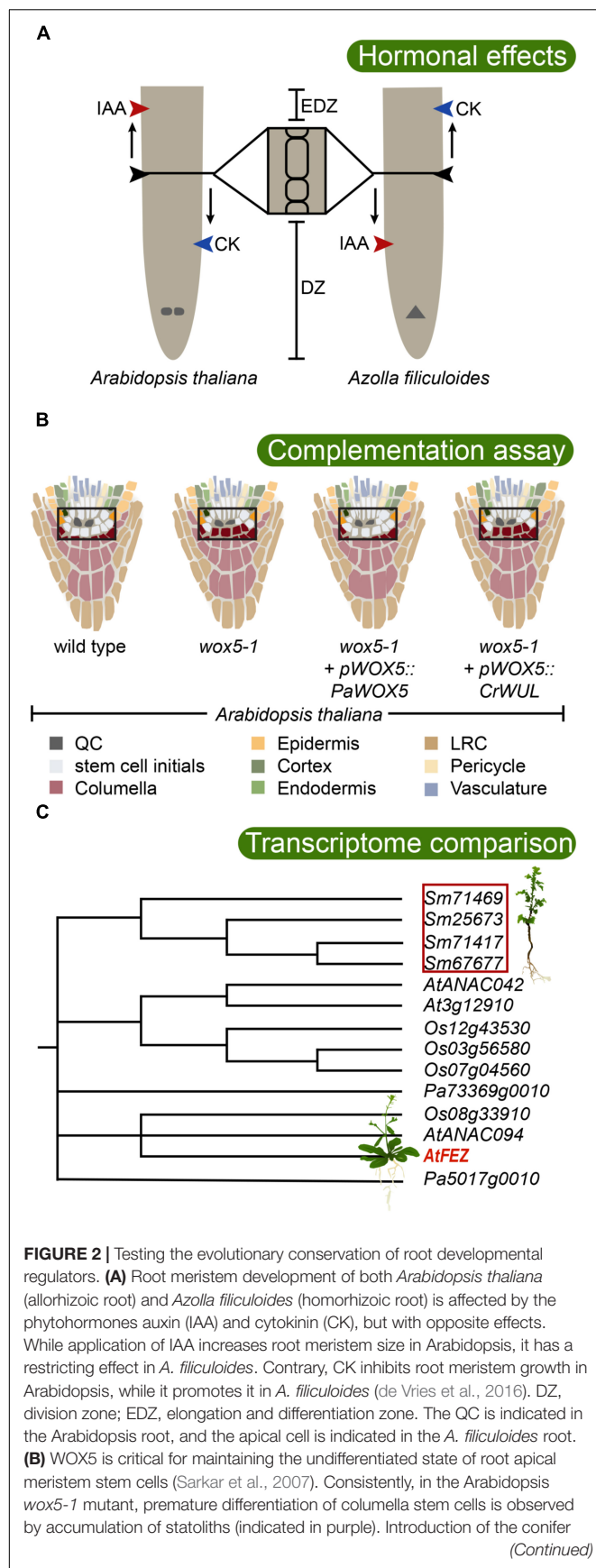
have occurred either by converting a shoot or by activating such a genetic circuit ensuring indeterminate growth elsewhere thereby triggering continuous growth *de novo*. Indeterminate growth is made possible by the activity of apical meristems that harbors pluripotent, constantly dividing cells. The activity of the meristem is ensured by stem cells (also called initial cells) that divide asymmetrically to give rise to cells that either undergo further divisions or begin to differentiate (Scheres, 2007). Stem cells for the specific tissue types of the root are found close to the root tip: distally (toward the tip) for the columella, proximally (shootward) for the stele and cortex/endodermis, and laterally distally for epidermis/lateral root cap (LRC). The proximal cells continue dividing within a division zone (DZ), until they reach a point in which division ceases and elongation begins. They then enter the elongation zone, and later the differentiation zone (collectively EDZ), where tissues are fully differentiated (**Figure 2A**; Dello Ioio et al., 2008; Perilli et al., 2012). The stem cells in the root apical meristem (RAM) are organized around mitotically inactive cells called quiescent center (QC) cells. The size of the QC (and the stem cell niche, i.e., QC + stem cells) varies substantially between species. *Arabidopsis* has a small meristem with only four QC cells. Within the monocots rice has 4–6, barley 30, and maize 500–1000 (Jiang et al., 2010; Kirschner et al., 2017). In a set of seminal cell laser ablation experiments, it was shown that if a QC cell in *Arabidopsis* is damaged, the neighboring columella stem cell begins to accumulate gravity sensing amyloplasts, showing that it is undergoing differentiation and suggesting that the QC sends a signal to surrounding stem cells to keep them undifferentiated (van den Berg et al., 1997).

Most of our current knowledge of the developmental genetic regulation of RAM establishment and maintenance comes from studies in *Arabidopsis*. Polar transport of auxin from the shoot down the root, *via* PIN auxin efflux carriers, creates an auxin maximum at the root tip which is needed for the maintenance of the stem cell niche. Here, auxin transport creates a “fountain” where auxin is refluxed up along the epidermis determining the location of the transition from DZ to EDZ (Blilou et al., 2005). In this process, auxin transport generates a concentration gradient over the meristem (Grieneisen et al., 2007). This auxin gradient determines the distribution of APETALA2-like PLETHORA (PLT) TFs that dose-dependently govern the extent of cell proliferation over the meristem (Galinha et al., 2007; Mähönen et al., 2014; Santuari et al., 2016). Considering the possibility that the RAM may have evolved from a pre-existing SAM it is interesting to note that the focused auxin maximum in the stem cells of the RAM is conceptually different from the SAM, where auxin maxima instead converge at the SAM periphery to promote lateral organ (e.g., leaf) formation (reviewed by Su et al., 2011). Auxin is transported away from these maxima and positive feedback from auxin concentration on auxin transport capacity canalizes auxin to narrow strands, which triggers procambium formation and xylem differentiation, while being transported toward the root. Within the SAM, cytokinin maintains cell division. In the root, however, cytokinin instead promotes cells to enter the EDZ and begin to differentiate. Therefore, a premature EDZ formation and smaller root meristem are the results of cytokinin application to roots (**Figure 2A**; Dello Ioio et al., 2008). Monocots

appear to respond in a similar manner to applied cytokinin as the root meristem of barley (*Hordeum vulgare*) becomes considerably smaller after application of cytokinin (Kirschner et al., 2018), suggesting conservation in hormonal regulation of root meristem size among angiosperms. Indeed, auxin and cytokinin-mediated regulation of plant development, including regulation of indeterminate growth, is strongly conserved and important also in bryophytes (Coudert et al., 2015; Flores-Sandoval et al., 2015; Bowman et al., 2017; Mutte et al., 2018; Thelander et al., 2018). Hence, this predates the evolution of vascular plants, and the evolution of roots, and we should therefore expect auxin and cytokinin to play important roles in lycophyte and fern root development.

In the water fern *Azolla filiculoides* which has a DZ and EDZ similar to an *Arabidopsis* root, application of cytokinin promotes cell division and enlarges the meristem, while auxin reduces it (**Figure 2A**; de Vries et al., 2016). Thus, the response in the *A. filiculoides* root to auxin and/or cytokinin is distinctly different from the response in *Arabidopsis*, and the fern root response resembles *Arabidopsis* SAM rather than RAM, with cytokinin promoting and auxin restricting meristem growth. These findings may support the hypothesis that euphyllophyte roots originated as postembryonically branching shoot structures, and that the seed plant primary root therefore is conceptually different and potentially non-homologous with the fern root (de Vries et al., 2016). Currently, it is not clear if the fern RAM requires an auxin maximum for its establishment and maintenance, but the inability of auxin to trigger lateral root formation in *Ceratopteris richardii*, which normally produces lateral roots from an internal endodermal cell layer (Hou et al., 2004), suggests that auxin plays different roles in fern roots compared to seed plant roots.

On the other hand, in the lycophyte *Selaginella kraussiana*, application of auxin promoted, while cytokinin inhibited, dichotomous branching of the roots, although roots forming after hormonal treatment were morphologically distorted (Sanders and Langdale, 2013). Also, many genes active in the *Selaginella moellendorffii* root are related to auxin, reinforcing the importance for this hormone in lycophyte root development (Huang and Schiefelbein, 2015). Intriguingly, analysis of fossils of an arborescent isoetalean lycophyte suggests that while the young plant had a shoot meristem and a “foot,” after the meristem had bifurcated, one of the two meristems bent downward forming a “rhizomorph” – a shoot with a rooting function – thereby generating a plant with an apparently bipolar organization (Sanders et al., 2010). Analysis of the patterns of xylem strands in the isoetalean fossil suggests that auxin was transported from the shoot down to the “root”-part (Sanders et al., 2010). Hence, reversion of polar auxin transport towards a positively gravitropic meristem could indicate a mechanism by which a shoot meristem may have been converted to a root meristem. This interpretation is supported by extant *S. kraussiana* clearly exhibiting basipetal auxin transport in shoots (Sanders and Langdale, 2013). Thus, there is support both for evolution of roots as converted shoots, and as entirely novel organs. One may therefore imagine that this occurred by several different type of mechanisms, enforcing the need to genetically and functionally assess root development and auxin response, in multiple lycophyte and fern species.

**FIGURE 2 |** Continued

Picea abies WOX5 homolog *PaWOX5*, driven by the WOX5 promoter, is able to restore the *wox5-1* phenotype, while the fern *Ceratopteris richardii* WUS/WOX5 homolog *CrWUL*, driven by the WOX5 promoter, cannot compensate for loss of WOX5 function (Zhang et al., 2017). This suggests that certain WOX5-specific properties evolved in the seed plant lineage. The stem cell niche is boxed, and colors inside of box are enhanced. **(C)** Transcriptome analysis revealed that a large number of angiosperm root meristem specific genes have homologs expressing in the root tip of the lycophyte *Selaginella moellendorffii* (Huang and Schiefelbein, 2015). In *Arabidopsis*, FEZ is required for LRC formation. The lycophyte root cap is an apparently analogous structure. Interestingly, *S. moellendorffii* expresses a set of FEZ-related genes in its roots suggesting the possibility of deep homology of factors regulating root cap development. The phylogenetic tree shows the well supported FEZ-clade, after Huang and Schiefelbein, 2015, including sequences from *Selaginella moellendorffii* (Sm), *Picea abies* (Pa), *Oryza sativa* (Os), and *Arabidopsis thaliana* (At). Poorly supported branchings are collapsed.

THE ROOT MERISTEM: A MIRROR IMAGE OF THE SHOOT MERISTEM?

Auxin and cytokinin as well as specific TFs are required for establishing and maintaining the RAM. Key TFs for maintenance of the QC and the stem cell niche are the PLTs (see above; Aida et al., 2004), the GRAS-type TFs SCARECROW (SCR) and SHORTROOT (SHR) (Sabatini et al., 2003), and the WUSCHEL-related homeobox5 (WOX5) (reviewed in Heyman et al., 2014). Interestingly, a recent study now link these factors, as both PLT and SCR were shown to interact with a third type of TF of the TCP type (Shimotohno et al., 2018). This PLT-SCR-TCP complex activates WOX5 expression, which is required for keeping stem cells, in particular columella stem cells undifferentiated. While PLT and SCR have broader activity domains, the WOX5 gene is specifically expressed in the QC (Sarkar et al., 2007). However, WOX5 acts non-cell autonomously because in the *wox5* mutant columella stem cells show signs of differentiation (Figure 2B; Sarkar et al., 2007). Hence, WOX5 may constitute the signal that was suggested by QC ablation experiments to emanate from the QC to maintain columella initial cells undifferentiated (van den Berg et al., 1997). The WOX5 protein has been shown to repress differentiation by forming a complex with a conserved repressor protein (Pi et al., 2015). WOX5 expression is restricted to the QC by peptide signaling from differentiated cells in the columella. This involves the CLE40 peptide which signals through the *Arabidopsis* CRINKLY4 (ACR4) receptor kinase (Stahl et al., 2009). This regulation is interesting from an evolutionary point of view because the WOX5 paralog WUSCHEL (WUS), which is active specifically in the organizing center of the SAM and required to maintain its stem cells, is similarly regulated by paralogous peptides and receptors (Stahl and Simon, 2010). Here, the CLV3 peptide signals from the outer L1 layer to restrict WUS to the central SAM a few cell layers below. Thus, if the RAM and the SAM are maintained by homologous factors, is this suggesting a common evolutionary origin for the two meristems?

While little is known of the evolutionary history of PLT and SCR type TFs, the WOX5/WUS genes have been studied more intensively. Despite being active in the two different meristems, the protein function of WOX5 and WUS has been conserved, as

shown by their interchangeability: WUS can restore a functional QC in a *wox5* mutant when directed by the WOX5 promoter-sequence, and WOX5 can compensate for loss of WUS in a similar type experiment (Sarkar et al., 2007). Tracing the phylogeny of WUS/WOX5 genes revealed orthologs of both WUS and WOX5 in both conifers and *Ginkgo*, but ferns have only one basal ortholog to these paralogous genes (Hedman et al., 2013). Thus, a gene duplication that gave rise to its co-orthologs WUS and WOX5 likely took place in the lineage leading to the seed plants. In *C. richardii*, the single ortholog *CrWUL* marks pluripotent cells in the shoot apex and also the proximal part of the root meristem, albeit not the root apical cell nor the distal side where the root cap is located (Nardmann and Werr, 2012). Assessing the ability of the fern and gymnosperm homologs to rescue the Arabidopsis *wus* or *wox5* mutant phenotypes, Zhang et al. (2017) found that both gymnosperm WUS and WOX5 homologs have conserved protein function, whereas the *CrWUL* protein was not able to replace neither WUS nor WOX5 (Figure 2B), unless it was specifically expressed in the columella initials. This shows that *CrWUL* has lost its cell-to-cell mobility, at least when expressed in Arabidopsis, but that it does have the possibility to interact in a similar molecular context and repress differentiation. In *C. richardii*, this is likely not relating to root cap development, as its activity domain is proximal to the apical cell of the meristem (Nardmann and Werr, 2012). Hence, this suggests that the central function for WUS/WOX5 factors to suppress differentiation is conserved in all euphyllophytes, but that special functions of WOX5 involving cell-to-cell movement between QC and columella have evolved in the seed plant lineage (Zhang et al., 2017). Thus, it is conceivable that following the polyploidization event that took place in the seed plant lineage (Jiao et al., 2011), a duplication of an already established genetic circuit allowed diverging in activity patterns, with one circuit regulating SAM and the other RAM maintenance. However, because *CrWUL* is active both in the fern RAM and SAM (Nardmann and Werr, 2012), the hypothesis, stating that the WOX5 circuit evolved from a SAM regulatory WUS circuit to control RAM, is maybe less likely. Instead, the WUS/WOX5 ability to promote stem cell maintenance under the control by peptide-mediated receptor kinase signaling with a potential for directional signaling could have been an efficient way of positioning a stem cell niche, that therefore would have been co-opted in different contexts during evolution. Indeed, WOX4, a more distant paralog still belonging to the same major clade as WUS/WOX5, similarly maintains the cambial stem cell niche. WOX4 is regulated by a CLE peptide (CLE41/TDIF) emanating from the phloem side of the cambium, sensed by a receptor kinase (PXY/TDR) in the cambium (Etchells et al., 2016). Thus, in common, there is a directional peptide signaling positioning the central region of a stem cell niche. The components for homologous peptide/receptor kinase signaling appear conserved in land plants (Nikonorova et al., 2015). It will be very interesting to know if *CrWUL* is similarly controlled by peptide/receptor kinase signaling, or if this evolved in the seed plant lineage.

In a recent opinion paper, Liu and Xu (2018) discuss the potential importance of a seed plant specific gene duplication of other paralogous WOX genes, belonging to the “intermediate

clade WOX” (IC-WOX; Hedman et al., 2013), for the allorhizoid root evolution in seed plants. IC-WOX genes are found in vascular plants (Liu and Xu, 2018). In the fern *C. richardii*, the IC-WOX homolog expresses specifically and transiently in root founder cells, during lateral and adventitious root development, suggesting that it is critical for root initiation (Nardmann and Werr, 2012). Similarly, its co-orthologs, *AtWOX11/12* expresses specifically in root founder cells during adventitious rooting (Hu and Xu, 2016; Liu et al., 2014). Intriguingly, the paralogs *AtWOX8/9* instead express specifically in the hypophyseal cell of the embryo, which gives rise to the QC and columella precursors of the seed plant primary/allorhizoid root (Breuninger et al., 2008). In both the adventitious roots and the embryonic root meristem WOX5 expression is initiated at a slightly later stage. Thus, Liu and Xu (2018) proposes that the gene duplication within the IC-WOX clade, which resulted in the *WOX11/12* genes that retained activity in adventitious root initiation and the *WOX8/9* genes that evolved a novel activity specifying an embryonic cell as a root founder cell, may have paved the way for a new type of root meristem giving rise to the allorhizoid/primary root.

THE ROOT CAP – PROTECTING, SENSING, AND SIGNALING

The RAM is maintained at the tip of the root, and as the root penetrates the soil there is an apparent risk of damaging this delicate structure. Hence, there most likely has been a strong selection force for the evolution of a structure protecting the RAM. Indeed, in all extant vascular plants, both euphyllophytes and lycophytes, the RAM is protected by a root cap (Kumpf and Nowack, 2015). The root cap consists of a columella and LRC. It facilitates the root's growth in soil due to production of mucilage, exudes various molecules, and may release long-lived cells into the rhizosphere to repel pathogens and attract symbionts (Kumpf and Nowack, 2015). Moreover, the root cap functions as a gravity-sensing organ rendering positive gravitropism to the root. For this purpose, the columella, in some plants also the LRC, harbors starch-containing amyloplasts, also called statoliths. The statoliths sediment with the gravity vector, which is sensed by the cell. In response, auxin flux in the RAM is modified, resulting in differential elongation of cells, and consequentially bending of the root tip with the gravity vector (Su et al., 2017). Furthermore, the LRC contributes hormonal cues that specifies cells competent for branching, thereby influencing how the entire root system may develop (Xuan et al., 2016). Thus, the root cap is a vital organ for the plant. This is consistent with the early and independent evolution of a root cap in lycophytes as well as in euphyllophytes.

There is quite a large variation in how the root cap is organized within root meristems of different species, and a number of different types have been described. The root cap can be clearly delimited from the QC and has its own stem cells as in Arabidopsis, which is therefore said to have a “closed” configuration. In other plants, such as in pea, the root cap initial cells are not clearly distinguishable from the QC and those meristems are said to be “open” (Kumpf and Nowack, 2015).

Gymnosperm meristems are open with no clear boundary between root cap initials and QC. Furthermore, in certain lycophytes, as in ferns, there is a single apical cell dividing in a distinct pattern to give rise to all cell types of the root, including the root cap, while in other lycophytes the meristem structure more resembles that of seed plants, and they can be either open or closed (Fujinami et al., 2017). A remarkably well preserved fossil of a progymnosperm with an active root meristem displays a clearly identifiable root cap surrounding a very broad meristem (Hetherington et al., 2016). Tracing domains of clonally related cells suggested that this carboniferous root meristem differed in organization from any previously described root meristem organization type. Hence, although extant plants display quite a variation in meristem/root cap organization an even greater diversity is likely to have existed in extinct plants. Intriguingly, all must be or must have been able to accommodate for a constant replenishing of their root caps.

As the root grows through the soil, the root cap cells are sloughed off or become released from the LRC *via* programmed cell death (Fendrych et al., 2014; Kumpf and Nowack, 2015). Hence, there is a need to produce new columella and LRC cells at the same rate as cells are lost. In *Arabidopsis*, the activity of *WOX5*, required for columella stem cell fate (Sarkar et al., 2007), is balanced by the activity of a TF of the NAC-family, called *FEZ*, which instead promotes the formation of LRC initials (Willemssen et al., 2008). *FEZ* also activates another NAC TF, called *SOMBRERO* (*SMB*), which together with *BEARSKIN1* and 2 promotes the differentiation of LRC cells. *SMB* in turn represses *FEZ* to prevent overproduction of LRC cells, while *WOX5* represses *SMB*, thereby controlling a precise development of the columella and LRC (Bennett et al., 2010, 2014). Despite the evolutionary importance of the root cap, surprisingly few studies exploring putative key genetic circuits shaping root cap development have been carried out in a comparative context. Recently, a transcriptome analysis of *S. moellendorffii* roots, comparing transcripts of the DZ and the EDZ, identified genes related to *FEZ* to be active in the roots of this lycophyte (Figure 2C; Huang and Schiefelbein, 2015). Detailed expression analyses and functional studies of these homologs may help us further understand the evolution of the root cap.

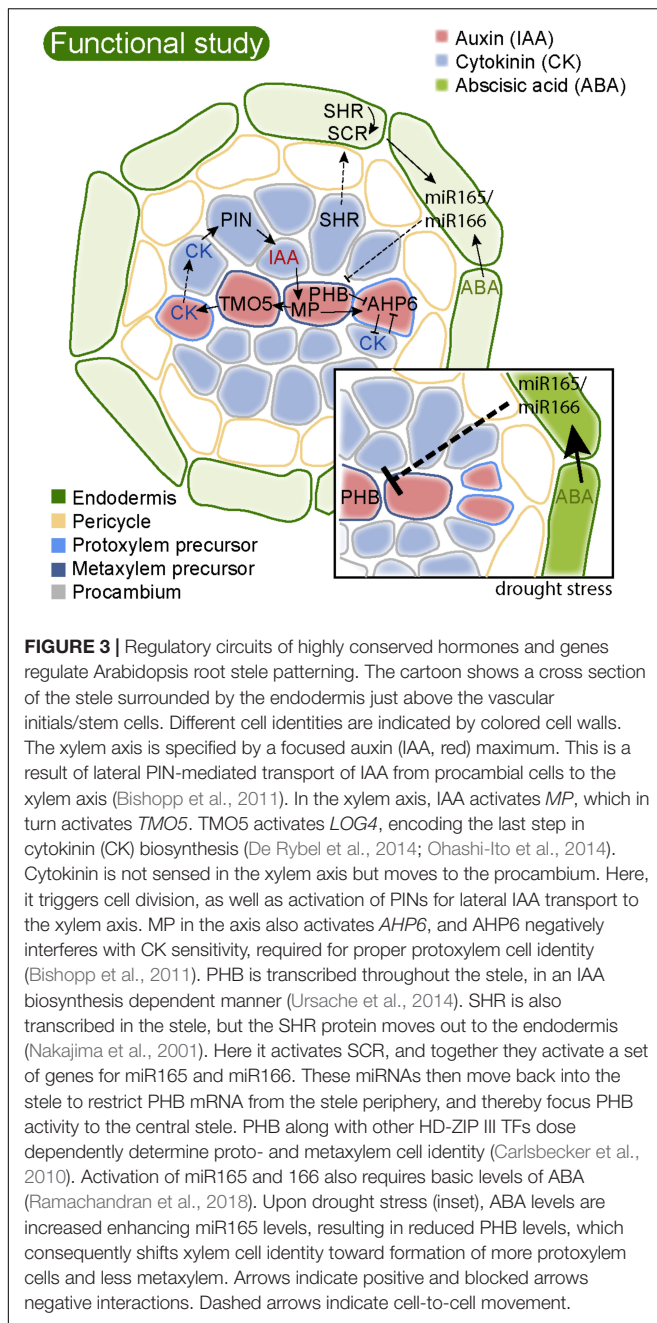
CELL-TO-CELL SIGNALING DETERMINES THE PATTERNING OF THE STELE

A vascular system with xylem and phloem that provides efficient transport of water, mineral nutrients, sugars, hormones, and other signaling molecules was a key evolutionary innovation, predating the evolution of true roots (Figure 1; Kenrick and Strullu-Derrien, 2014; Taylor et al., 2009). As the root meristem generates cells within the DZ, these cells acquire specific identities, primarily depending on their position relative to each other (Yu et al., 2017). The procambium, the meristematic tissue from which the primary vascular tissues are derived, is localized in the center surrounded by the ground and dermal tissues. Within the procambium xylem and phloem precursor

cells are patterned in a species-specific manner, in most roots in a protostele arrangement. In a protostele, xylem forms in the center and may be arching out in distinct patterns, flanked by procambium and phloem. This pattern is established already in the embryo and propagated by the RAM (De Rybel et al., 2016). In *Arabidopsis*, the stele has a diarch arrangement, i.e., an axis of xylem traverses the stele. Within the stele, two types of primary xylem vessels are formed: protoxylem with spiral or annular secondary cell walls, and metaxylem with reticulated or pitted walls. Protoxylem develops at the periphery of the axis, while metaxylem later differentiates at the center of the stele. This is an exarch pattern. Also in lycophyte roots, a protostele is common, but here the xylem pattern is opposite to most other plants, and protoxylem forms in the center and metaxylem toward the periphery (endarch pattern). In certain lycophyte roots, the endarch vascular tissues instead surrounds a pith, in a siphonostele arrangement. In monocots, multiple xylem strands may surround a pith, but as in other angiosperms, the xylem forms in an exarch pattern, with metaxylem toward the pith and protoxylem toward endodermis (Taylor et al., 2009).

The vascular pattern in *Arabidopsis* is determined by a number of evolutionarily conserved factors, including hormones and TFs. Auxin synthesis, transport, and signaling are required for the establishment of the central stele (De Rybel et al., 2016). Auxin is needed to break the initial radial symmetry of the root with the formation of a xylem axis (De Rybel et al., 2014; Bishopp et al., 2011). This process is reinforced by the antagonistic action of cytokinin, which in turn is required for cell proliferation in the neighboring procambium. High auxin levels triggers activation of the AUXIN RESPONSE FACTOR 5 (*ARF5*)/*MONOPTEROS* (*MP*). *MP* in turn activates *TARGET OF MONOPTEROS5* (*TMO5*) in the xylem axis (Schlereth et al., 2010). *TMO5* then directly activates cytokinin biosynthesis (De Rybel et al., 2014; Ohashi-Ito et al., 2014). The presumably high cytokinin level is not sensed within the xylem axis, but instead in the neighboring procambial cells, to which it diffuses or is transported. In the procambium, cytokinin promotes cell division, but it also promotes auxin transporters (*PIN*s) that move auxin toward the xylem axis (Bishopp et al., 2011). The high auxin level in the xylem axis also activates a specific cytokinin signaling component, *AHP6*, which instead of transmitting the cytokinin signal acts to inhibit it (Bishopp et al., 2011). Hence, in *Arabidopsis* there is a mutually inhibitory action of auxin and cytokinin which is defining the root vascular pattern (Figure 3). If any of these components are disturbed, this will alter the pattern of xylem and procambium in the stele; in particular, the protoxylem is sensitive to perturbations of auxin and cytokinin (Bishopp et al., 2011).

Patterning of the root xylem axis with peripheral protoxylem and central metaxylem requires TFs of the class III homeo-domain leucine zipper (*HD-ZIP III*) family, regulated by *miR165* and *miR166* (Carlsbecker et al., 2010; Miyashima et al., 2011, reviewed in Ramachandran et al., 2017). The *HD-ZIP III* genes as well as *miR166* have homologs in all land plants (Floyd and Bowman, 2004, 2006; Prigge and Clark, 2006). In moss gametophytes, *HD-ZIP III* TFs regulate leaf development



(Yip et al., 2016). Because bryophytes are non-vascular plants, these factors must have been recruited to regulate vascular development during the evolution of vascular tissues. Indeed, HD-ZIP III expression is detected in vascular tissues in lycopphytes (Floyd and Bowman, 2006; Prigge and Clark, 2006). In Arabidopsis, the HD-ZIP III family includes five members, and mutant phenotypes suggest that they dose dependently specify the xylem cell type identity. Plants lacking all five HD-ZIP III transcription factors fail to develop xylem (Carlsbecker et al., 2010). The miR165/166 regulating HD-ZIP III in the root are produced in the endodermis where they are activated by SHR together with its paralog SCR. SHR is produced in the stele

but moves out to the endodermis to activate SCR (Helariutta et al., 2000). Together, these TFs induce the expression of genes coding for miR165 and miR166, which in turn move back into the stele. At the peripheral stele, high levels of miR165 and miR166 strongly reduces the abundance of HD-ZIP III mRNA, in particular of PHABULOSA (PHB). The relatively low HD-ZIP III protein level determines protoxylem cell identity, while less miR165/166 in the center allow high levels of HD-ZIP III TFs governing metaxylem formation (**Figure 3**; Carlsbecker et al., 2010; Miyashima et al., 2011). The HD-ZIP III TFs are tightly interlinked with auxin and cytokinin signaling. All HD-ZIP III genes are directly or indirectly requiring auxin biosynthesis for their transcriptional activation (Ursache et al., 2014), and in turn they modulate both auxin and cytokinin signaling and synthesis components (Carlsbecker et al., 2010; Dello Ioio et al., 2012; Müller et al., 2016). Because of the complex interactions between hormones, TFs and small RNAs, mathematical modeling have been employed to assess what components and parameters are required for reaching a protostele pattern with a traversing xylem axis with peripheral protoxylem and central metaxylem (Mellor et al., 2017).

Although the vascular pattern is inherently distinctive for distinct species, it also appears to be plastic within a species to some degree, allowing endogenous and external cues to modify the pattern. Abiotic stress, such as drought, results in a vascular pattern with extra protoxylem strands flanking the central metaxylem in Arabidopsis (Ramachandran et al., 2018). Drought stress is mediated by the hormone abscisic acid, ABA, and ABA applications result in a similar pattern as drought. This pattern is similar to lower order HD-ZIP III mutants, and indeed, elevated ABA cause increases in miR165, resulting in reduced HD-ZIP III levels (**Figure 3**, inset; Ramachandran et al., 2018). The levels of miR165 and miR166 have been found to vary with external conditions in a variety of species (Zhao et al., 2007; Liu et al., 2008; Ding et al., 2009) suggesting that modulation of their activity has the potential to change developmental patterning also in these species. It will be very interesting to see if these factors and the auxin/cytokinin balance may underlie the distinct vascular patterning of various species. Among the angiosperms the monocots display a rather different arrangement, with a siphonostele. Importantly, there was a recent report of fluorescent auxin and cytokinin signaling reporters in barley plants, allowing live tracking of hormonal signaling (Kirschner et al., 2018). Such an approach will be needed to understand if and how auxin and cytokinin pattern also the monocot root vasculature. It is obvious that polar auxin transport plays an important role in vascular development in extant plants, but fossil evidence also suggest it was important in now extinct plants. Stuningly, analyses of wood of fossil plants of both arborescent lycopphytes and progymnosperms reveal a circular pattern of treachery elements above buds and branch junctions in stems (Rothwell et al., 2009). Such patterns are also seen in extant trees and emanates from routes of polar auxin transport. Thus, this provides evidence for polar auxin transport in vascular tissue formation in 375 million years old lycopphytes, and suggests that canalization of auxin was coupled to the evolution of vasculature tissues (Rothwell et al., 2009). Recently, Zhu et al. analyzed the

transcriptome of *S. moellendorffii* stems and found that many key factors, such as SHR/SCR, HD-ZIP III, and TMO5, have homologs in this lycophyte although certain components of the gene regulatory network required for Arabidopsis root vascular patterning were not identified. Thus, the regulatory mechanisms of lycophyte vascular development is perhaps less complex or involves different components than in flowering plants (Zhu et al., 2017). Modeling may generate hypotheses for how patterns such as the siphonostele, or the endarch protostele pattern of lycophytes are established. Such hypotheses may be tested by mapping gene expression and regulatory networks in lycophytes and other phylogenetically informative species.

THE STELE ENSURES IT IS SURROUNDED BY A SINGLE GUARDING ENDODERMAL LAYER

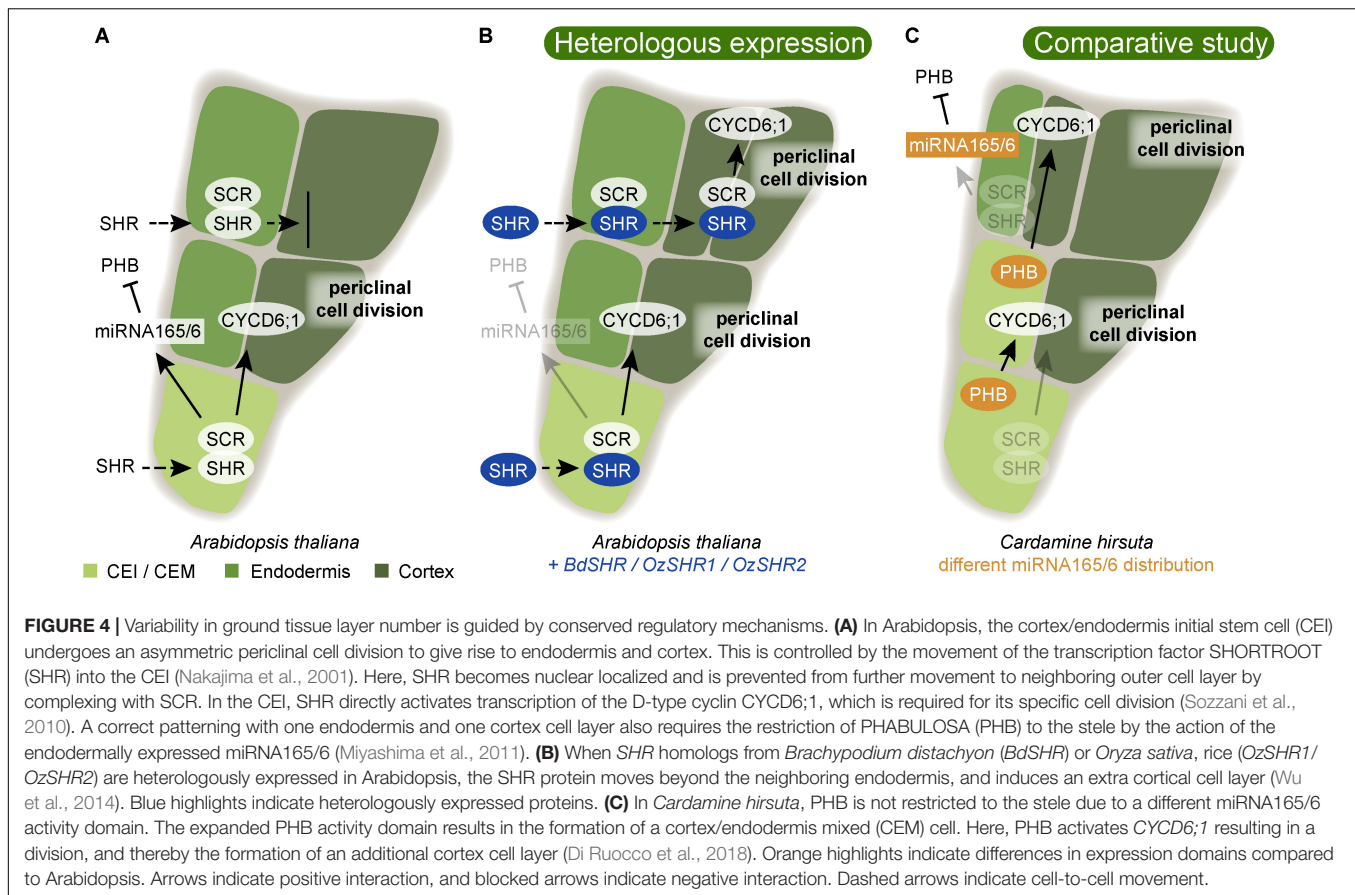
The central function for the root is to take up water and mineral nutrients. In this process, the endodermis forms an apoplastic barrier with the Casparian strip and suberin lamellae restricting diffusion of water nutrients, and thereby allowing a controlled uptake (for reviews, see Geldner, 2013; Barberon and Geldner, 2014). This very specialized cell layer likely evolved at least twice, in the lycophytes and in the euphyllophytes. Fossils of early species of each of these two lineages apparently lacked an endodermal layer, and extant *Lycopodium* does not have a root endodermal layer (Raven and Edwards, 2001; Kenrick and Strullu-Derrien, 2014; Raven, 2018). Hence, the endodermis may have evolved as a relatively recent innovation in each lineage (Figure 1). The endodermis is the inner layer of the ground tissue of the root, which outside of the endodermis harbors the cortex. The cortex generally consists of parenchymatic cells and can provide several different functions such as storage or, by the formation of aerenchyma, means to improve flooding tolerance. The outer layer of the cortex may also develop an exodermis, a first barrier inside the epidermis (Kim et al., 2018).

In all plants that have an endodermis, there is only one layer, just next to the stele. Thus, genetic mechanisms have to operate to delimit the specific differentiation to only this ground tissue layer. The prevailing hypothesis of how plants ensure the formation of a single endodermal layer just outside the stele relies on molecular communication from the stele providing both positional information and information for endodermal differentiation (Wu et al., 2014; Doblas et al., 2017b). In Arabidopsis, both ground tissue layers, cortex and endodermis, originate from the same cortex/endodermis initial stem cell (CEI) which first divides anticlinally. The daughter cell then undergoes an asymmetric periclinal division to produce one endodermal and one cortical cell layer (Figure 4A). If either SHR or SCR is mutated, this periclinal division does not occur, and only one ground tissue layer is formed (Di Laurenzio et al., 1996; Helariutta et al., 2000). In the *shr* mutant, this layer exhibits cortex characteristics suggesting that SHR is required for endodermis differentiation (Helariutta et al., 2000), while in the *scr* mutant, the single layer exhibits a mix between cortex and endodermis characteristics (Di Laurenzio et al., 1996). Important

for the positioning of the endodermis just outside of the stele is the movement of SHR into the neighboring outer cell layer from the stele, where it is expressed (Nakajima et al., 2001). Here, SHR activates SCR, with which it forms a complex in the nucleus (Cui et al., 2007; Welch et al., 2007). This prevents SHR from moving further and prevents additional periclinal cell divisions thereby ensuring the formation of only one endodermal layer. Together, SHR and SCR trigger the asymmetric periclinal cell division resulting in the endodermal and cortex layers by direct induction of *cyclin D6;1* (*CYCD6;1*) (Figure 4A; Sozzani et al., 2010).

Wu et al. (2014) tested the potentially conserved functions of monocot SHR by introducing SHR homologs from *Brachypodium distachyon*, *BdSHR*, and *Oryza sativa* (rice), *OsSHR1*, and *OsSHR2*, into Arabidopsis (Figure 4B). As expected, both *BdSHR* and *OsSHR1/2* were able to activate and bind to Arabidopsis SCR. However, the movement of the SHR homologs was not restricted to one layer, but they continued moving, triggering the formation of additional cortex, but not endodermal layers (Figure 4B; Wu et al., 2014). Thus, this finding may uncover a potentially important role for SHR/SCR to trigger multiple cortex divisions. It is likely that this is an important mechanism in monocots that often have many cortex layers. This experiment also revealed that SHR alone is not sufficient to induce endodermis differentiation. Instead, additional conserved signals from the stele together with SHR are likely required for determination of a single endodermal layer. SHR and SCR are highly conserved. In conifers, the *Pinus sylvestris* (*PsySCR*) homolog is specifically expressed in the endodermis and the ground tissue initials (Laajanen et al., 2007) and also SHR homologs have been found in conifer roots (Solé et al., 2008). Going even further back the land plant phylogeny, Zhu et al. (2017) found SHR and SCR homologs in the transcriptome of *S. moellendorffii* roots, stems, and leaves. However, the presence of homologous genes might be a good first indication but does not necessarily mean that the function is conserved as well. For instance, SCR and SHR homologs were also found to be essential for bundle sheath specification in leaves (Cui et al., 2014; Yoon et al., 2016). Thus, more detailed analyses are required to test the hypothesis that these homologs perform similar functions as their Arabidopsis homologs.

The differentiation of the endodermis involves the formation of a Casparian strip, specific depositions of lignin in the cell wall between the endodermal cells, providing an apoplastic barrier. Next step is incorporation of a suberin-containing lamellae in the wall, while specific cells are passage cells and are kept open for intake of molecules (Geldner, 2013; Doblas et al., 2017a; Andersen et al., 2018). In this process, SHR acts at the top of a gene regulatory cascade, and directly activates another key TF, MYB36, and these two TFs activate genes for both Casparian strip and suberin lamellae differentiation (Kamiya et al., 2015; Liberman et al., 2015). Interestingly, a stele derived peptide, which signals into the endodermal layer, ensures proper maintenance of the Casparian strip, providing additional molecular surveillance from the stele on the endodermis (Doblas et al., 2017b). Recently, a study found that the genetic regulation



of endodermis formation is highly conserved in tomato (*Solanum lycopersicon*) (Li et al., 2018), but the genetic regulation of endodermis differentiation might be conserved also outside of angiosperms. Indeed, phylogenetic analyses could identify homologs to many other key factors in all plants with an endodermis (Li et al., 2018). However, although CASP proteins, responsible for localization of the Casparian strip, are highly conserved among plants, only euphyllophytes have CASPs with a specific protein domain important for their function (Roppolo et al., 2014). Continued research into evolutionary aspects of the components now rapidly being discovered in *Arabidopsis* promises to shed light on endodermis evolution within a close future.

While the function of the endodermis as a barrier for water and nutrient uptake is well established, the purpose of varying amounts of cortex layers is less obvious. Upon the observation that cortex proliferation can be induced by oxidative stress, Cui (2015) speculated that cortex proliferation might be a protective mechanism against abiotic stress. On the other hand, there is evidence for a trend in plant evolution to produce thinner roots, presumably to improve the efficiency of soil exploration and to reduce the dependence on symbiotic mycorrhiza (Ma et al., 2018). Accordingly, a reduced root cortical cell file number in maize was correlating with improved drought tolerance (Chimungu et al., 2014). A recent study identified a mechanism for generating multiple cortex layers in

Cardamine hirsuta, a close relative of *Arabidopsis*. Di Ruocco et al. (2018) showed that levels of miR165/166 not only are important for vascular development (see above) but also for the determination of cortical cell number in *C. hirsuta*. Enhanced *MIR165A* expression causes the formation of only one instead of two cortical cell layer, while reduced miR165/6 activity in *Arabidopsis* caused additional formation of cortex layers by an expanded expression domain of PHB triggering CYCD6;1 activation, similar to SHR/SCR (Figure 4C; Di Ruocco et al., 2018). Thus, small changes in miRNA activity can have a big impact on root anatomy and may underlie anatomical differences between species.

OUTLOOK

As we have seen, we have quite a detailed understanding on how morphogenesis and anatomy of the *Arabidopsis* root is established. How the broadly similar morphology and anatomy of distantly related monocot, gymnosperm, fern, or lycophyte roots are genetically controlled is, however, largely unknown. A better understanding of the underlying genetic regulation will allow us to view the evolution of roots in a clearer light. Although roots are essential for almost all vascular plants, for agriculture, and for ecosystems we have a rather limited understanding of how this essential organ has evolved, but also how its development

is regulated in most species. Hence, despite the importance of roots, there are quite a few outstanding questions remaining to be answered. Has the root evolved as a modified shoot, as the presence of homologous regulatory factors may suggest. Or is the root an entirely novel organ, as the opposite auxin transport patterns in the shoot and root meristem indicates. Is the primary, allorhizous, seed plant root homologous with the adventitious, homorhizous, roots of ferns? Is there “deep homology” as potentially indicated by the identification of putative homologs to key root development regulators in lycophytes? How could complex and central structures for root function such as the root cap and the endodermis have evolved independently both in lycophyte and euphyllophyte roots? How did the intricate cell-to-cell communication required for root patterning evolve?

Addressing these and other questions will be facilitated by the very rapid technology development and data generation from next-generation sequencing approaches. Current efforts in characterizing transcriptome and genome sequences of the lycophyte *S. moellendorffii* (Banks et al., 2011), several fern species, including *C. richardii* (Wolf et al., 2015), the conifer *Picea abies* (Nystedt et al., 2013) in addition to the vast amount of data that is accumulating for non-vascular “outgroup” plants, such as the moss *Physcomitrella patens* (Rensing et al., 2008) and the liverwort *Marchantia polymorpha* (Bowman et al., 2017), are providing information on genetic advances that occurred during land plant’s first evolutionary steps as well as when the seed plants evolved, and beyond. Initiatives to sequence the genomes of yet a large number of phylogenetically important vascular plants, both non-seed plant and seed plants, within the 10K initiative, which is leveraging the 1K effort of sequencing 1000 plant genomes (Cheng et al., 2018), will most likely substantially contribute to illuminating various aspects of how roots may have evolved.

At the same time as we are exploring the vast diversity among species and their morphologies and anatomies, it will be important to develop non-angiosperm models of vascular plants (Schulz et al., 2010). Models allow building of knowledge within a research community, for detailed comparative studies with non-model plants by various approaches. It will be essential to establish protocols for transformation of plants to allow reverse genetics. Currently, there is an efficient transformation protocol for the fern *C. richardii* (Plackett et al., 2014), *A. filiculoides* is emerging as another rapidly growing fern model with great potential (de Vries and de Vries, 2018), several species of *Selaginella* are emerging lycopod models (Schulz et al., 2010), and transformation protocols and various resources exist for the conifer *P. abies* (Uddenberg et al., 2015). Furthermore, in a model species detailed gene expression analyses using laser capture microdissection coupled to RNA sequencing, or even single cell approaches, are feasible, and will provide opportunities

to build detailed gene expression maps. This will be instrumental for co-expression analyses and construction of gene regulatory networks. Such networks can be compared with the detailed gene regulatory network around key developmental regulators in *Arabidopsis* (Taylor-Teeple et al., 2014; Santuari et al., 2016; Drapek et al., 2017) to allow inferences of important shifts potentially underlying evolutionary novelties. Together with hormone signaling localization and detailed morphological and anatomical studies of potential changes resulting from external signaling or perturbation, it would allow inferring developmental core modules responsible for specific features. In such a system meaningful heterologous complementation experiments can be conducted with key genes from closely or distantly related species, to test conservation of protein function. In **Figures 2–4** we point out various approaches by which knowledge of a process in the model plant *Arabidopsis* can be used to widen our understanding of similar processes in other plants. With established fern and lycopod models we can extend this type of analyses substantially. Along with transcriptome data from a dense phylogenetic sampling, we are on the way to a comprehensive understanding of the underlying genetic key factors for morphological features such as the RAM, root cap, endodermis, or specific stele patterns. Mirroring morphological and anatomical outcomes of genetic and hormonal perturbation experiments with the phenotypes of extant, but also extinct fossil morphologies and anatomies, will allow us to formulate specific and testable hypotheses on how genetic networks may be rewired during evolution to generate novel morphologies, or even novel organs – such as the repeated evolution of roots. There are indeed exciting times ahead when we dig deeper into the evolution and developmental biology of plant roots.

AUTHOR CONTRIBUTIONS

FA and AC prepared and finalized the manuscript together.

FUNDING

Research on conifer root and vascular development was funded by the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (FORMAS; 2013-953 and 2017-00857) to AC.

ACKNOWLEDGMENTS

The authors thank Daniel Uddenberg, Jan de Vries, Peter Engström, and Prashanth Ramachandran for thoughtful discussions.

REFERENCES

- Ahmed, M. A., Zarebanadkouki, M., Meunier, F., Javaux, M., Kaestner, A., and Carminati, A. (2018). Root type matters: measurement of water uptake by seminal, crown, and lateral roots in maize. *J. Exp. Bot.* 69, 1199–1206. doi: 10.1093/jxb/erx439
- Aida, M., Beis, D., Heidstra, R., Willemsen, V., Blilou, I., Galinha, C., et al. (2004). The PLETHORA genes mediate patterning of the *Arabidopsis*

- root stem cell niche. *Cell* 119, 109–120. doi: 10.1016/j.cell.2004.09.018
- Andersen, T. G., Naseer, S., Ursache, R., Wybouw, B., Smet, W., De Rybel, B., et al. (2018). Diffusible repression of cytokinin signalling produces endodermal symmetry and passage cells. *Nature* 555, 529–533. doi: 10.1038/nature25976
- 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
- Barberon, M., and Geldner, N. (2014). Radial transport of nutrients: the plant root as a polarized epithelium. *Plant Physiol.* 166, 528–537. doi: 10.1104/pp.114.246124
- Bennett, T., van den Toorn, A., Sanchez-Perez, G. F., Campilho, A., Willemsen, V., Snel, B., et al. (2010). SOMBRERO, BEARSKIN1, and BEARSKIN2 regulate root cap maturation in *Arabidopsis*. *Plant Cell* 22, 640–654. doi: 10.1105/tpc.109.072272
- Bennett, T., van den Toorn, A., Willemsen, V., and Scheres, B. (2014). Precise control of plant stem cell activity through parallel regulatory inputs. *Development* 141, 4055–4064. doi: 10.1242/dev.110148
- Bishopp, A., Help, H., El-Showk, S., Weijers, D., Scheres, B., Friml, J., et al. (2011). A mutually inhibitory interaction between auxin and cytokinin specifies vascular pattern in roots. *Curr. Biol.* 21, 917–926. doi: 10.1016/j.cub.2011.04.017
- Blilou, I., Xu, J., Wildwater, M., Willemsen, V., Paponov, I., Friml, J., et al. (2005). The PIN auxin efflux facilitator network controls growth and patterning in *Arabidopsis* roots. *Nature* 433, 39–44. doi: 10.1038/nature03184
- 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.e15. doi: 10.1016/j.cell.2017.09.030
- Breuninger, H., Rikirsch, E., Hermann, M., Ueda, M., and Laux, T. (2008). Differential expression of WOX genes mediates apical-basal axis formation in the *Arabidopsis* embryo. *Dev. Cell* 14, 867–876. doi: 10.1016/j.devcel.2008.03.008
- Carlsbecker, A., Lee, J.-Y., Roberts, C. J., Dettmer, J., Lehesranta, S., Zhou, J., et al. (2010). Cell signalling by microRNA165/6 directs gene dose-dependent root cell fate. *Nature* 465, 316–321. doi: 10.1038/nature08977
- Cheng, S., Melkonian, M., Smith, S. A., Brockington, S., Archibald, J. M., Delaux, P.-M., et al. (2018). 10KP: a phylodiverse genome sequencing plan. *Gigascience* 7, 1–9. doi: 10.1093/gigascience/giy013
- Chimungu, J. G., Brown, K. M., and Lynch, J. P. (2014). Reduced root cortical cell file number improves drought tolerance in maize. *Plant Physiol.* 166, 1943–1955. doi: 10.1104/pp.114.249037
- Coudert, Y., Palubicki, W., Ljung, K., Novák, O., Leyser, O., and Harrison, C. J. (2015). Three ancient hormonal cues co-ordinate shoot branching in a moss. *eLife* 4:e06808. doi: 10.7554/eLife.06808
- Cui, H. (2015). Cortex proliferation in the root is a protective mechanism against abiotic stress. *Plant Signal. Behav.* 10:e1011949. doi: 10.1080/15592324.2015.1011949
- Cui, H., Kong, D., Liu, X., and Hao, Y. (2014). SCARECROW, SCR-LIKE 23 and SHORT-ROOT control bundle sheath cell fate and function in *Arabidopsis thaliana*. *Plant J.* 78, 319–327. doi: 10.1111/tpj.12470
- Cui, H., Levesque, M. P., Vernoux, T., Jung, J. W., Paquette, A. J., Gallagher, K. L., et al. (2007). An evolutionarily conserved mechanism delimiting SHR movement defines a single layer of endodermis in plants. *Science* 316, 421–425. doi: 10.1126/science.1139531
- De Rybel, B., Adibi, M., Breda, A. S., Wendrich, J. R., Smit, M. E., Novák, O., et al. (2014). Plant development. Integration of growth and patterning during vascular tissue formation in *Arabidopsis*. *Science* 345:1255215. doi: 10.1126/science.1255215
- De Rybel, B., Mähönen, A. P., Helariutta, Y., and Weijers, D. (2016). Plant vascular development: from early specification to differentiation. *Nat. Rev. Mol. Cell Biol.* 17, 30–40. doi: 10.1038/nrm.2015.6
- de Vries, J., Fischer, A. M., Roettger, M., Rommel, S., Schluepmann, H., Bräutigam, A., et al. (2016). Cytokinin-induced promotion of root meristem size in the fern *Azolla* supports a shoot-like origin of euphyllphyte roots. *New Phytol.* 209, 705–720. doi: 10.1111/nph.13630
- de Vries, S., and de Vries, J. (2018). “*Azolla*: model system for symbiotic nitrogen fixation and evolutionary developmental biology,” in *Current Advances in Fern Research*, ed. H. Fernández (Cham: Springer), 21–46. doi: 10.1007/978-3-319-75103-0_2
- Dello Ioio, R., Galinha, C., Fletcher, A. G., Grigg, S. P., Molnar, A., Willemsen, V., et al. (2012). A PHABULOSA/cytokinin feedback loop controls root growth in *Arabidopsis*. *Curr. Biol.* 22, 1699–1704. doi: 10.1016/j.cub.2012.07.005
- Dello Ioio, R., Nakamura, K., Moubayidin, L., Perilli, S., Taniguchi, M., Morita, M. T., et al. (2008). A genetic framework for the control of cell division and differentiation in the root meristem. *Science* 322, 1380–1384. doi: 10.1126/science.1164147
- Di Laurenzio, L., Wysocka-Diller, J., Malamy, J. E., Pysh, L., Helariutta, Y., Freshour, G., et al. (1996). The SCARECROW gene regulates an asymmetric cell division that is essential for generating the radial organization of the *Arabidopsis* root. *Cell* 86, 423–433. doi: 10.1016/S0092-8674(00)80115-4
- Di Ruocco, G., Bertolotti, G., Pacifici, E., Polverari, L., Tsiantis, M., Sabatini, S., et al. (2018). Differential spatial distribution of miR165/6 determines variability in plant root anatomy. *Development* 145:dev153858. doi: 10.1242/dev.153858
- Ding, D., Zhang, L., Wang, H., Liu, Z., Zhang, Z., and Zheng, Y. (2009). Differential expression of miRNAs in response to salt stress in maize roots. *Ann. Bot.* 103, 29–38. doi: 10.1093/aob/mcn205
- Doblas, V. G., Geldner, N., and Barberon, M. (2017a). The endodermis, a tightly controlled barrier for nutrients. *Curr. Opin. Plant Biol.* 39, 136–143. doi: 10.1016/j.pbi.2017.06.010
- Doblas, V. G., Smakowska-Luzan, E., Fujita, S., Alassimone, J., Barberon, M., Madalinski, M., et al. (2017b). Root diffusion barrier control by a vasculature-derived peptide binding to the SGN3 receptor. *Science* 355, 280–284. doi: 10.1126/science.aaj1562
- Drapek, C., Sparks, E. E., and Benfey, P. N. (2017). Uncovering gene regulatory networks controlling plant cell differentiation. *Trends Genet.* 33, 529–539. doi: 10.1016/j.tig.2017.05.002
- Etchells, J. P., Smit, M. E., Gaudinier, A., Williams, C. J., and Brady, S. M. (2016). A brief history of the TDIF-PXY signalling module: balancing meristem identity and differentiation during vascular development. *New Phytol.* 209, 474–484. doi: 10.1111/nph.13642
- Fendrych, M., Van Hautegeem, T., Van Durme, M., Olvera-Carrillo, Y., Huysmans, M., Karimi, M., et al. (2014). Programmed cell death controlled by ANAC033/SOMBRERO determines root cap organ size in *Arabidopsis*. *Curr. Biol.* 24, 931–940. doi: 10.1016/j.cub.2014.03.025
- 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
- Floyd, S. K., and Bowman, J. L. (2004). Gene regulation: ancient microRNA target sequences in plants. *Nature* 428, 485–486. doi: 10.1038/428485a
- Floyd, S. K., and Bowman, J. L. (2006). Distinct developmental mechanisms reflect the independent origins of leaves in vascular plants. *Curr. Biol.* 16, 1911–1917. doi: 10.1016/j.cub.2006.07.067
- Fujinami, R., Yamada, T., Nakajima, A., Takagi, S., Idogawa, A., Kawakami, E., et al. (2017). Root apical meristem diversity in extant lycophytes and implications for root origins. *New Phytol.* 215, 1210–1220. doi: 10.1111/nph.14630
- Galinha, C., Hofhuis, H., Luijten, M., Willemsen, V., Blilou, I., Heidstra, R., et al. (2007). PLETHORA proteins as dose-dependent master regulators of *Arabidopsis* root development. *Nature* 449, 1053–1057. doi: 10.1038/nature06206
- Geldner, N. (2013). The Endodermis. *Annu. Rev. Plant Biol.* 64, 531–558. doi: 10.1146/annurev-arplant-050312-120050
- Grieneisen, V. A., Xu, J., Marée, A. F. M., Hogeweg, P., and Scheres, B. (2007). Auxin transport is sufficient to generate a maximum and gradient guiding root growth. *Nature* 449, 1008–1013. doi: 10.1038/nature06215
- Hedman, H., Zhu, T., von Arnold, S., and Sohlberg, J. J. (2013). Analysis of the WUSCHEL-RELATED HOMEBOX gene family in the conifer *Picea abies* reveals extensive conservation as well as dynamic patterns. *BMC Plant Biol.* 13:89. doi: 10.1186/1471-2229-13-89

- Helariutta, Y., Fukaki, H., Wysocka-Diller, J., Nakajima, K., Jung, J., Sena, G., et al. (2000). The SHORT-ROOT gene controls radial patterning of the *Arabidopsis* root through radial signaling. *Cell* 101, 555–567. doi: 10.1016/S0092-8674(00)80865-X
- Hetherington, A. J., and Dolan, L. (2017). The evolution of lycopsid rooting structures: conservatism and disparity. *New Phytol.* 215, 538–544. doi: 10.1111/nph.14324
- Hetherington, A. J., and Dolan, L. (2018a). Bilaterally symmetric axes with rhizoids composed the rooting structure of the common ancestor of vascular plants. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 373:20170042. doi: 10.1098/rstb.2017.0042
- Hetherington, A. J., and Dolan, L. (2018b). Stepwise and independent origins of roots among land plants. *Nature* 52:381. doi: 10.1038/s41586-018-0445-z
- Hetherington, A. J., Dubrovsky, J. G., and Dolan, L. (2016). Unique cellular organization in the oldest root meristem. *Curr. Biol.* 26, 1629–1633. doi: 10.1016/j.cub.2016.04.072
- Heyman, J., Kumpf, R. P., and De Veylder, L. (2014). A quiescent path to plant longevity. *Trends Cell Biol.* 24, 443–448. doi: 10.1016/j.tcb.2014.03.004
- Honkanen, S., and Dolan, L. (2016). Growth regulation in tip-growing cells that develop on the epidermis. *Curr. Opin. Plant Biol.* 34, 77–83. doi: 10.1016/j.pbi.2016.10.006
- Honkanen, S., Jones, V. A. S., Morieri, G., Champion, C., Hetherington, A. J., Kelly, S., et al. (2017). The mechanism forming the cell surface of tip-growing rooting cells is conserved among land plants. *Curr. Biol.* 27, 3238–3224. doi: 10.1016/j.cub.2017.09.062
- Hou, G., Hill, J. P., and Blancaflor, E. B. (2004). Developmental anatomy and auxin response of lateral root formation in *Ceratopteris richardii*. *J. Exp. Bot.* 55, 685–693. doi: 10.1093/jxb/erh068
- Hu, X., and Xu, L. (2016). Transcription factors WOX11/12 directly activate WOX5/7 to promote root primordia initiation and organogenesis. *Plant Physiol.* 172, 2363–2373. doi: 10.1104/pp.16.01067
- Huang, L., and Schiefelbein, J. (2015). Conserved gene expression programs in developing roots from diverse plants. *Plant Cell* 27, 2119–2132. doi: 10.1105/tpc.15.00328
- Jiang, K., Zhu, T., Diao, Z., Huang, H., and Feldman, L. J. (2010). The maize root stem cell niche: a partnership between two sister cell populations. *Planta* 231, 411–424. doi: 10.1007/s00425-009-1059-3
- Jiao, Y., Wickett, N. J., Ayyampalayam, S., Chanderbali, A. S., Landherr, L., Ralph, P. E., et al. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature* 473, 97–100. doi: 10.1038/nature09916
- Jones, V. A. S., and Dolan, L. (2012). The evolution of root hairs and rhizoids. *Ann. Bot.* 110, 205–212. doi: 10.1093/aob/mcs136
- Kamiya, T., Borghi, M., Wang, P., Danku, J. M. C., Kalmbach, L., Hosmani, P. S., et al. (2015). The MYB36 transcription factor orchestrates Casparian strip formation. *Proc. Natl. Acad. Sci. U.S.A.* 112, 10533–10538. doi: 10.1073/pnas.1507691112
- Kenrick, P., and Strullu-Derrien, C. (2014). The origin and early evolution of roots. *Plant Physiol.* 166, 570–580. doi: 10.1104/pp.114.244517
- Kim, Y. X., Ranathunge, K., Lee, S., Lee, Y., Lee, D., and Sung, J. (2018). Composite transport model and water and solute transport across plant roots: an update. *Front. Plant Sci.* 9:193. doi: 10.3389/fpls.2018.00193
- Kirschner, G. K., Stahl, Y., Imani, J., von Korff, M., and Simon, R. (2018). Fluorescent reporter lines for auxin and cytokinin signalling in barley (*Hordeum vulgare*). *PLoS One* 13:e0196086. doi: 10.1371/journal.pone.0196086
- Kirschner, G. K., Stahl, Y., von Korff, M., and Simon, R. (2017). Unique and conserved features of the barley root meristem. *Front. Plant Sci.* 8:1240. doi: 10.3389/fpls.2017.01240
- Kumpf, R. P., and Nowack, M. K. (2015). The root cap: a short story of life and death. *J. Exp. Bot.* 66, 5651–5662. doi: 10.1093/jxb/erv295
- Laajanen, K., Vuorinen, I., Salo, V., Juuti, J., and Raudaskoski, M. (2007). Cloning of *Pinus sylvestris* SCARECROW gene and its expression pattern in the pine root system, mycorrhiza and NPA-treated short roots. *New Phytol.* 175, 230–243. doi: 10.1111/j.1469-8137.2007.02102.x
- Li, P., Yang, M., Chang, J., Wu, J., Zhong, F., Rahman, A., et al. (2018). Spatial expression and functional analysis of Casparian strip regulatory genes in endodermis reveals the conserved mechanism in tomato. *Front. Plant Sci.* 9:832. doi: 10.3389/fpls.2018.00832
- Liberman, L. M., Sparks, E. E., Moreno-Risueno, M. A., Petricka, J. J., and Benfey, P. N. (2015). MYB36 regulates the transition from proliferation to differentiation in the *Arabidopsis* root. *Proc. Natl. Acad. Sci. U.S.A.* 112, 12099–12104. doi: 10.1073/pnas.1515576112
- Liu, H.-H., Tian, X., Li, Y.-J., Wu, C.-A., and Zheng, C.-C. (2008). Microarray-based analysis of stress-regulated microRNAs in *Arabidopsis thaliana*. *RNA* 14, 836–843. doi: 10.1261/rna.895308
- Liu, J., Sheng, L., Xu, Y., Li, J., Yang, Z., Huang, H., et al. (2014). WOX11 and 12 are involved in the first-step cell fate transition during de novo root organogenesis in *Arabidopsis*. *Plant Cell* 26, 1081–1093. doi: 10.1105/tpc.114.1.22887
- Liu, W., and Xu, L. (2018). Recruitment of IC-WOX genes in root evolution. *Trends Plant Sci.* 23, 490–496. doi: 10.1016/j.tplants.2018.03.011
- Ma, Z., Guo, D., Xu, X., Lu, M., Bardgett, R. D., Eissenstat, D. M., et al. (2018). Evolutionary history resolves global organization of root functional traits. *Nature* 555, 94–97. doi: 10.1038/nature25783
- Mähönen, A. P., Ten Tusscher, K., Siligato, R., Smetana, O., Díaz-Triviño, S., Salojärvi, J., et al. (2014). PLETHORA gradient formation mechanism separates auxin responses. *Nature* 515, 125–129. doi: 10.1038/nature13663
- Matsunaga, K. K. S., and Tomescu, A. M. F. (2016). Root evolution at the base of the lycophyte clade: insights from an Early Devonian lycophyte. *Ann. Bot.* 117, 585–598. doi: 10.1093/aob/mcw006
- Mellor, N., Adibi, M., El-Showk, S., De Rybel, B., King, J., Mähönen, A. P., et al. (2017). Theoretical approaches to understanding root vascular patterning: a consensus between recent models. *J. Exp. Bot.* 68, 5–16. doi: 10.1093/jxb/erw410
- Menand, B., Yi, K., Jouannic, S., Hoffmann, L., Ryan, E., Linstead, P., et al. (2007). An ancient mechanism controls the development of cells with a rooting function in land plants. *Science* 316, 1477–1480. doi: 10.1126/science.1142618
- Miyashima, S., Koi, S., Hashimoto, T., and Nakajima, K. (2011). Non-cell-autonomous microRNA165 acts in a dose-dependent manner to regulate multiple differentiation status in the *Arabidopsis* root. *Development* 138, 2303–2313. doi: 10.1242/dev.060491
- Müller, C. J., Valdés, A. E., Wang, G., Ramachandran, P., Beste, L., Uddenberg, D., et al. (2016). PHABULOSA mediates an auxin signaling loop to regulate vascular patterning in *Arabidopsis*. *Plant Physiol.* 170, 956–970. doi: 10.1104/pp.15.01204
- Mutte, S. K., Kato, H., Rothfels, C., Melkonian, M., Wong, G. K.-S., and Weijers, D. (2018). Origin and evolution of the nuclear auxin response system. *eLife* 7, e33399. doi: 10.7554/eLife.33399
- Nakajima, K., Sena, G., Nawy, T., and Benfey, P. N. (2001). Intercellular movement of the putative transcription factor SHR in root patterning. *Nature* 413, 307–311. doi: 10.1038/35095061
- Nardmann, J., and Werr, W. (2012). The invention of WUS-like stem cell-promoting functions in plants predates leptosporangiate ferns. *Plant Mol. Biol.* 78, 123–134. doi: 10.1007/s11103-011-9851-4
- Nikonorova, N., Vu, L. D., Czyzewicz, N., Gevaert, K., and De Smet, I. (2015). A phylogenetic approach to study the origin and evolution of the CRINKLY4 family. *Front. Plant Sci.* 6:880. doi: 10.3389/fpls.2015.00880
- Nystedt, B., Street, N. R., Wetterbom, A., Zuccolo, A., Lin, Y.-C., Scofield, D. G., et al. (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature* 497, 579–584. doi: 10.1038/nature12211
- Ohashi-Ito, K., Saegusa, M., Iwamoto, K., Oda, Y., Katayama, H., Kojima, M., et al. (2014). A bHLH complex activates vascular cell division via cytokinin action in root apical meristem. *Curr. Biol.* 24, 2053–2058. doi: 10.1016/j.cub.2014.07.050
- Perilli, S., Di Mambro, R., and Sabatini, S. (2012). Growth and development of the root apical meristem. *Curr. Opin. Plant Biol.* 15, 17–23. doi: 10.1016/j.pbi.2011.10.006
- Pi, L., Aichinger, E., van der Graaff, E., Llavata-Peris, C. I., Weijers, D., Hennig, L., et al. (2015). Organizer-derived WOX5 signal maintains root columella stem cells through chromatin-mediated repression of CDF4 expression. *Dev. Cell* 33, 576–588. doi: 10.1016/j.devcel.2015.04.024
- Pires, N. D., and Dolan, L. (2012). Morphological evolution in land plants: new designs with old genes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 508–518. doi: 10.1098/rstb.2011.0252
- Plackett, A. R. G., Huang, L., Sanders, H. L., and Langdale, J. A. (2014). High-efficiency stable transformation of the model fern species *Ceratopteris richardii*

- via microparticle bombardment. *Plant Physiol.* 165, 3–14. doi: 10.1104/pp.113.231357
- Prigge, M. J., and Clark, S. E. (2006). Evolution of the class III HD-Zip gene family in land plants. *Evol. Dev.* 8, 350–361. doi: 10.1111/j.1525-142X.2006.00107.x
- Puttick, M. N., Morris, J. N., 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.e2. doi: 10.1016/j.cub.2018.01.063
- Ramachandran, P., Carlsbecker, A., and Etchells, J. P. (2017). Class III HD-ZIPs govern vascular cell fate: an HD view on patterning and differentiation. *J. Exp. Bot.* 68, 55–69. doi: 10.1093/jxb/erw370
- Ramachandran, P., Wang, G., Augstein, F., de Vries, J., and Carlsbecker, A. (2018). Continuous root xylem formation and vascular acclimation to water deficit involves endodermal ABA signalling via miR165. *Development* 145:dev159202. doi: 10.1242/dev.159202
- Raven, J. A. (2018). Evolution and palaeophysiology of the vascular system and other means of long-distance transport. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 373:20160497. doi: 10.1098/rstb.2016.0497
- Raven, J. A., and Edwards, D. (2001). Roots: evolutionary origins and biogeochemical significance. *J. Exp. Bot.* 52, 381–401. doi: 10.1093/jxb/52.suppl_1.381
- Rensing, S. A., Lang, D., Zimmer, A. D., Terry, A., Salamov, A., Shapiro, H., et al. (2008). The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* 319, 64–69. doi: 10.1126/science.1150646
- Roppolo, D., Boeckmann, B., Pfister, A., Boutet, E., Rubio, M. C., Dénervaud-Tendon, V., et al. (2014). Functional and evolutionary analysis of the casparian strip membrane domain protein family. *Plant Physiol.* 165, 1709–1722. doi: 10.1104/pp.114.239137
- Rothwell, G. W., Sanders, H., Wyatt, S. E., and Lev-Yadun, S. (2009). A fossil record for growth regulation: the role of auxin in wood evolution. *Ann. Mo Bot. Gard.* 95, 121–134. doi: 10.3417/2006208
- Sabatini, S., Heidstra, R., Wildwater, M., and Scheres, B. (2003). SCARECROW is involved in positioning the stem cell niche in the *Arabidopsis* root meristem. *Genes Dev.* 17, 354–358. doi: 10.1101/gad.252503
- Sanders, H., Rothwell, G. W., and Wyatt, S. E. (2010). Parallel evolution of auxin regulation in rooting systems. *Plant Syst. Evol.* 291, 221–225. doi: 10.1007/s00606-010-0383-1
- Sanders, H. L., and Langdale, J. A. (2013). Conserved transport mechanisms but distinct auxin responses govern shoot patterning in *Selaginella kraussiana*. *New Phytol.* 198, 419–428. doi: 10.1111/nph.12183
- Santuari, L., Sanchez-Perez, G., Luijten, M., Rutjens, B., Terpstra, I., Berke, L., et al. (2016). The PLETHORA gene regulatory network guides growth and cell differentiation in *Arabidopsis* roots. *Plant Cell* 28, 2937–2951. doi: 10.1105/tpc.16.00656
- Sarkar, A. K., Luijten, M., Miyashima, S., Lenhard, M., Hashimoto, T., Nakajima, K., et al. (2007). Conserved factors regulate signalling in *Arabidopsis thaliana* shoot and root stem cell organizers. *Nature* 446, 811–814. doi: 10.1038/nature05703
- Scheres, B. (2007). Stem-cell niches: nursery rhymes across kingdoms. *Nat. Rev. Mol. Cell Biol.* 8, 345–354. doi: 10.1038/nrm2164
- Schlereth, A., Möller, B., Liu, W., Kientz, M., Flipse, J., Rademacher, E. H., et al. (2010). MONOPTEROS controls embryonic root initiation by regulating a mobile transcription factor. *Nature* 464, 913–916. doi: 10.1038/nature08836
- Schulz, C., Little, D. P., Stevenson, D. W., Nowogrodzki, A., and Paquiot, D. (2010). Growth and care instructions of a new model species—the lycophyte *Selaginella apoda*. *Am. Fern J.* 100, 167–171. doi: 10.1640/0002-8444-100.3.167
- Scotland, R. W. (2010). Deep homology: a view from systematics. *Bioessays* 32, 438–449. doi: 10.1002/bies.200900175
- Shimotohno, A., Heidstra, R., Blilou, I., and Scheres, B. (2018). Root stem cell niche organizer specification by molecular convergence of PLETHORA and SCARECROW transcription factor modules. *Genes Dev.* 32, 1085–1100. doi: 10.1101/gad.314096.118
- Solé, A., Sánchez, C., Vielba, J. M., Valladares, S., Abarca, D., and Díaz-Sala, C. (2008). Characterization and expression of a *Pinus radiata* putative ortholog to the *Arabidopsis* SHORT-ROOT gene. *Tree Physiol.* 28, 1629–1639. doi: 10.1093/treephys/28.11.1629
- Sozzani, R., Cui, H., Moreno-Risueno, M. A., Busch, W., Van Norman, J. M., Vernoux, T., et al. (2010). Spatiotemporal regulation of cell-cycle genes by SHORTROOT links patterning and growth. *Nature* 466, 128–132. doi: 10.1038/nature09143
- Stahl, Y., and Simon, R. (2010). Plant primary meristems: shared functions and regulatory mechanisms. *Curr. Opin. Plant Biol.* 13, 53–58. doi: 10.1016/j.pbi.2009.09.008
- Stahl, Y., Wink, R. H., Ingram, G. C., and Simon, R. (2009). A signaling module controlling the stem cell niche in *Arabidopsis* root meristems. *Curr. Biol.* 19, 909–914. doi: 10.1016/j.cub.2009.03.060
- Su, S.-H., Gibbs, N. M., Jancewicz, A. L., and Masson, P. H. (2017). Molecular mechanisms of root gravitropism. *Curr. Biol.* 27, R964–R972. doi: 10.1016/j.cub.2017.07.015
- Su, Y.-H., Liu, Y.-B., and Zhang, X.-S. (2011). Auxin-cytokinin interaction regulates meristem development. *Mol. Plant* 4, 616–625. doi: 10.1093/mp/ssr007
- Tam, T. H. Y., Catarino, B., and Dolan, L. (2015). Conserved regulatory mechanism controls the development of cells with rooting functions in land plants. *Proc. Natl. Acad. Sci. U.S.A.* 112, E3959–E3968. doi: 10.1073/pnas.1416324112
- Taylor, T. N., Taylor, E. L., and Krings, M. (2009). *Paleobotany: The biology and Evolution of Fossil Plants*, 2nd Edn. New York, NY: Academic Press.
- Taylor-Teeple, M., Lin, L., de Lucas, M., Turco, G., Toal, T. W., Gaudinier, A., et al. (2014). An *Arabidopsis* gene regulatory network for secondary cell wall synthesis. *Nature* 517, 571–575. doi: 10.1038/nature14099
- Thelander, M., Landberg, K., and Sundberg, E. (2018). Auxin-mediated developmental control in the moss *Physcomitrella patens*. *J. Exp. Bot.* 69, 277–290. doi: 10.1093/jxb/erx255
- Uddenberg, D., Akhter, S., Ramachandran, P., Sundström, J. F., and Carlsbecker, A. (2015). Sequenced genomes and rapidly emerging technologies pave the way for conifer evolutionary developmental biology. *Front. Plant Sci.* 6:970. doi: 10.3389/fpls.2015.00970
- Ursache, R., Miyashima, S., Chen, Q., Vaten, A., Nakajima, K., Carlsbecker, A., et al. (2014). Tryptophan-dependent auxin biosynthesis is required for HD-ZIP III-mediated xylem patterning. *Development* 141, 1250–1259. doi: 10.1242/dev.103473
- van den Berg, C., Willemsen, V., Hendriks, G., Weisbeek, P., and Scheres, B. (1997). Short-range control of cell differentiation in the *Arabidopsis* root meristem. *Nature* 390, 287–289. doi: 10.1038/36856
- Welch, D., Hassan, H., Blilou, I., Immink, R., Heidstra, R., and Scheres, B. (2007). *Arabidopsis* JACKDAW and MAGPIE zinc finger proteins delimit asymmetric cell division and stabilize tissue boundaries by restricting SHORT-ROOT action. *Genes Dev.* 21, 2196–2204. doi: 10.1101/gad.440307
- Willemsen, V., Bauch, M., Bennett, T., Campilho, A., Wolkenfelt, H., Xu, J., et al. (2008). The NAC domain transcription factors FEZ and SOMBRERO control the orientation of cell division plane in *Arabidopsis* root stem cells. *Dev. Cell* 15, 913–922. doi: 10.1016/j.devcel.2008.09.019
- Wolf, P. G., Sessa, E. B., Marchant, D. B., Li, F.-W., Rothfels, C. J., Sigel, E. M., et al. (2015). An exploration into fern genome space. *Genome Biol. Evol.* 7, 2533–2544. doi: 10.1093/gbe/evv163
- Wu, S., Lee, C.-M., Hayashi, T., Price, S., Divol, F., Henry, S., et al. (2014). A plausible mechanism, based upon SHORT-ROOT movement, for regulating the number of cortex cell layers in roots. *Proc. Natl. Acad. Sci. U.S.A.* 111, 16184–16189. doi: 10.1073/pnas.1407371111
- Xuan, W., Band, L. R., Kumpf, R. P., Van Damme, D., Parizot, B., De Rop, G., et al. (2016). Cyclic programmed cell death stimulates hormone signaling and root development in *Arabidopsis*. *Science* 351, 384–387. doi: 10.1126/science.aad2776
- Yip, H. K., Floyd, S. K., Sakakibara, K., and Bowman, J. L. (2016). Class III HD-Zip activity coordinates leaf development in *Physcomitrella patens*. *Dev. Biol.* 419, 184–197. doi: 10.1016/j.ydbio.2016.01.012
- Yoon, E. K., Dhar, S., Lee, M.-H., Song, J. H., Lee, S. A., Kim, G., et al. (2016). Conservation and diversification of the SHR-SCR-SCL23 regulatory network in the development of the functional endodermis in *Arabidopsis* shoots. *Mol. Plant* 9, 1197–1209. doi: 10.1016/j.molp.2016.06.007
- Yu, Q., Li, P., Liang, N., Wang, H., Xu, M., and Wu, S. (2017). Cell-fate specification in *Arabidopsis* roots requires coordinative action of lineage instruction and positional reprogramming. *Plant Physiol.* 175, 816–827. doi: 10.1104/pp.17.00814

- Zhang, Y., Jiao, Y., Jiao, H., Zhao, H., and Zhu, Y.-X. (2017). Two-step functional innovation of the stem-cell factors WUS/WOX5 during plant evolution. *Mol. Biol. Evol.* 34, 640–653. doi: 10.1093/molbev/msw263
- Zhao, B., Liang, R., Ge, L., Li, W., Xiao, H., Lin, H., et al. (2007). Identification of drought-induced microRNAs in rice. *Biochem. Biophys. Res. Commun.* 354, 585–590. doi: 10.1016/j.bbrc.2007.01.022
- Zhu, Y., Chen, L., Zhang, C., Hao, P., Jing, X., and Li, X. (2017). Global transcriptome analysis reveals extensive gene remodeling, alternative splicing and differential transcription profiles in non-seed vascular plant *Selaginella moellendorffii*. *BMC Genomics* 18:1042. doi: 10.1186/s12864-016-3266-1

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 © 2018 Augstein and Carlsbecker. 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.



Activation of Nucleases, PCD, and Mobilization of Reserves in the *Araucaria angustifolia* Megagametophyte During Germination

Laura Moyano^{1,2†}, María D. Correa^{1†}, Leonardo C. Favre^{3,4}, Florencia S. Rodríguez¹, Sara Maldonado^{1,2*} and María P. López-Fernández^{1,2}

¹ Departamento de Biodiversidad y Biología Experimental, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina, ² Consejo Nacional de Investigaciones Científicas Técnicas, Instituto de Biodiversidad y Biología Experimental y Aplicada, Buenos Aires, Argentina, ³ Departamentos de Industrias y Departamento de Química Orgánica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina, ⁴ Consejo Nacional de Investigaciones Científicas y Técnicas, Buenos Aires, Argentina

OPEN ACCESS

Edited by:

Verónica S. Di Stilio,
University of Washington,
United States

Reviewed by:

Andrew Leslie,
Brown University, United States
Rosemary White,
Commonwealth Scientific
and Industrial Research Organisation
(CSIRO), Australia

*Correspondence:

Sara Maldonado
saram@bg.fcen.uba.ar

[†] These authors have contributed
equally to this work

Specialty section:

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

Received: 10 May 2018

Accepted: 14 August 2018

Published: 30 August 2018

Citation:

Moyano L, Correa MD, Favre LC,
Rodríguez FS, Maldonado S and
López-Fernández MP (2018)
Activation of Nucleases, PCD,
and Mobilization of Reserves
in the *Araucaria angustifolia*
Megagametophyte During
Germination. *Front. Plant Sci.* 9:1275.
doi: 10.3389/fpls.2018.01275

The megagametophyte of mature seeds of *Araucaria angustifolia* consists of cells with thin walls, one or more nuclei, a central vacuole storing proteins, and a cytoplasm rich in amyloplasts, mitochondria and lipid bodies. In this study, we describe the process of mobilization of reserves and analyzed the dismantling of the tissue during germination, using a range of well-established markers of programmed cell death (PCD), including: morphological changes in nuclei and amyloplasts, DNA degradation, and changes in nuclease profiles. TUNEL reaction and DNA electrophoresis demonstrate that DNA fragmentation in nuclei occurs at early stages of germination, which correlates with induction of specific nucleases. The results of the present study add knowledge on the dismantling of the megagametophyte of genus *Araucaria*, a storage tissue that stores starch as the main reserve substance, as well as on the PCD pathway, by revealing new insights into the role of nucleases and the expression patterns of putative nuclease genes during germination.

Keywords: *Araucaria angustifolia*, PCD, nucleases, starch, Cys-EP, megagametophyte, germination

INTRODUCTION

In seeds of both Gymnosperms and Angiosperms, stored nutrients must be mobilized to support germination and early seedling growth (Young and Gallie, 2000a; Bewley et al., 2013). During germination, the main seed storage tissues, i.e., the endosperm, the perisperm, or both (in Angiosperms), and the megagametophyte (in Gymnosperms), undergo programmed cell death (PCD). In Angiosperms, the mobilization of lipids and proteins from lipid and protein bodies during germination has been studied in seeds of several species. In cereal seeds, for example, it is well documented that cells of the aleurone layer lack a central vacuole and store proteins and lipids in protein vacuoles and lipid bodies, respectively. It is also known that vacuole fusion is necessary for the establishment of the large central vacuole, which is the site where various hydrolytic enzymes and other molecules involved in PCD are localized (Zheng et al., 2017), and that

damages to the integrity of the tonoplast alter the integrity of the plasma membrane, causing the collapse and subsequent death of the cell (Bethke et al., 1999). A similar process has been described in the endosperm of Dicots species such as tomato (Bewley et al., 2013), *Datura ferox* (Mella et al., 1995), and castor bean (Gietl and Schmid, 2001). However, in Gymnosperms, to date, the cell death of the megagametophyte during the mobilization of reserves is understudied and the process has been described only in *Araucaria bidwillii* (Casani et al., 2009) and *Picea glauca* (He and Kermode, 2003).

The megagametophyte of mature seeds of *Araucaria angustifolia* consists of cells with thin walls, one or more nuclei, a cytoplasm that is rich in amyloplasts, mitochondria and lipid bodies and a central vacuole that stores proteins (Panza et al., 2002). Starch is the most conspicuous reserve (Panza et al., 2002). The high water content characterizing *A. angustifolia* mature seeds (ca. 40%) is contained in the large central vacuole. Tompsett (1984) examined the relationship between seed moisture content and germination after desiccation in nine *Araucaria* species and established three moisture content groups: a group composed of *A. araucana*, *A. angustifolia*, *A. hunsteinii*, and *A. bidwillii*, which cannot be dried to below 25–40% without damage; a second group composed of *A. columnaris*, *A. rulei*, *A. nemorosa*, and *A. scopulorum*, which cannot be dried to below 12% without damage; and a third group composed of *A. cunninghamii*, which can be dried to 2% without damage. This author also found that seeds of the first group are larger and heavier and are mainly starchy, whereas those in the other groups possess mainly lipid content, and are smaller and lighter. Starchy seeds are found in species of a major clade of *Araucaria* that includes the extant sections *Araucana*, *Bunya*, and *Intermedia*, whereas oily seeds are found in species of the section *Eutacta* (Setoguchi et al., 1998).

Several reports have shown that cells undergoing PCD show the presence of some nucleases (deoxyribonucleases and ribonucleases) (Sakamoto and Takami, 2014 and references therein). To date, various plant deoxyribonucleases have been reported. Of these, several endonucleases and an exonuclease in *Arabidopsis* seem to act in leaf senescence because they were shown to be inducible at the transcript level (Sakamoto and Takami, 2014). During germination, endonucleases have been identified in the aleurone layer of cereals (Young and Gallie, 1999; Fath et al., 2000, among others) and in the embryo axes from French bean (Lambert et al., 2014).

In addition to endonucleases, KDEL-tailed cysteine endopeptidases (Cys-EPs), a group of papain-type peptidases, have been found in senescing tissue. These peptidases are synthesized as proenzymes with a C-terminal KDEL endoplasmic reticulum retention signal (Schmid et al., 1998). The signal is removed, and the enzyme separates from the endoplasmic reticulum in small vesicles called ricinosomes. Cys-EPs are able to digest extensins, which are the proteins that form the basic support for the structure of the cell wall. Cys-EPs have been detected in the endosperm of *Ricinus communis* (Schmid et al., 2001), the epigeal cotyledons of *Vigna mungo* (Toyooka et al., 2001), and the megagametophyte of *Picea glauca* (He and Kermode, 2010) during germination, as well as in the micropylar

endosperm and suspensor of *Chenopodium quinoa* during seed development (López-Fernández and Maldonado, 2013b).

In the present study, we assessed the PCD of the megagametophyte of *A. angustifolia* seeds during germination, with the objective to evaluate the expression and activity of nucleases in cells that reserve starch, as well as the sequence of autophagy and PCD during the process of mobilization of reserves. After analyzing the previous reports mentioned above, we inferred that the PCD pathway of the megagametophyte of *A. angustifolia* is different from that of the aleurone layer in cereals, since, in the former, the central vacuole already exists and the main reserve is located in the plastids. The PCD pathway in *A. angustifolia* should also be different from that of the starchy endosperm of cereals (Young and Gallie, 1999, 2000b; Sabelli, 2012; Domínguez and Cejudo, 2014) and that of the starchy perisperm of quinoa (López-Fernández and Maldonado, 2013a; Burrieza et al., 2014), since, in these tissues, PCD occurs during the development of the seed and is associated with the accumulation and not with the dismantling of the reserves. It should be clarified that, in Gymnosperms, the cell death of the megagametophyte during the mobilization of reserves has been investigated in *Araucaria bidwillii* (Casani et al., 2009), a species that, like *A. angustifolia*, also produces starchy seeds. In this species, necrosis (and probably also PCD in some cells) was identified by DNA fragmentation, changes in the size and morphology of nuclei, and a substantial increase in proteolytic activities, including those of caspase-like proteases. It is also worth mentioning that, in *Araucaria araucana*, starch degradation is initiated by α -amylase and phosphorylase in the embryo and by phosphorylase mainly in the megagametophyte (Cardemil and Reinero, 1982; Cardemil and Varner, 1984).

To describe the PCD of the megagametophyte during germination of *A. angustifolia* seeds, in the present study we analyzed the mobilization of reserves at different times following imbibition, and investigated the characteristics that define the process of PCD and autophagy such as activation of Cys-EPs, nuclear fragmentation and internucleosomal DNA cleavage. Likewise, we analyzed genes of S1 nuclease-like endonucleases and *Staphylococcus* nuclease-like (SN) endonucleases and a gene with a DNase-RNase domain not classified as S1 or as Tudor because it lacks these domains.

MATERIALS AND METHODS

Plant Material

Araucaria angustifolia seeds were collected from trees grown in natural populations in the Botanical Garden “Arturo E. Ragonese”-INTA Castelar, situated in Buenos Aires province, Argentina (34°40'S 58°39'W), from March to May 2017. Seeds were surface-disinfected with 5% NaClO for 15 min and then allowed to germinate onto imbibed perlite in a growth chamber under controlled conditions of 16 h light/8 h dark cycles at 25°C. At 14, 28, and 42 days after germination (DAG), specifically following radicle protrusion, the seeds were dissected, and the megagametophytes were either used fresh or milled after freeze-drying and the flours stored at −80°C until use.

Experiments reported here were repeated with at least three independent biological replicates; the results were comparable across experiments, unless otherwise stated.

Sample Preparation for Histological Analysis

Samples were collected at 0, 14, 28, and 42 DAG and prepared for microscopy according to López-Fernández and Maldonado (2013a) by fixation in 4% paraformaldehyde, 0.1 M phosphate buffered saline (PBS) pH 7.2 for 24 h at 4°C. After rinsing, the samples were dehydrated in an acetone series, and then embedded in Technovit 8100 (Kulzer and Co., Germany). Resin was polymerized at 4°C. The sections were stained with 0.5% toluidine blue O (Sigma-Aldrich, St. Louis, MO, United States) in aqueous solution, or used without staining procedure.

For the TUNEL assay, samples were fixed at 4°C in 4% paraformaldehyde (0.1 PBS; pH 7.2), dehydrated in a graded ethanol series (30, 40, 50, 60, 70, 80, 90, and 100%) and embedded in LRW resin (Polyscience, Inc., Warrington, PA, United States; 17411) as previously described by Harris et al. (1995). Semi-thin sections (1 µm thick) were mounted on glass slides. To identify starch and proteins, sections were stained with Lugol solution (Biopack 151205, Argentina) and Amido Black (Anedra 6952, Argentina), respectively (Jensen, 1962; Owens et al., 1993).

Evans Blue Staining

Megagametophytes at 0 and 42 DAG following germination were stained with 1% Evans Blue for 1 min, destained with deionized water for 1 h, and photographed under a dissecting microscope.

RNA Extraction and Semi-Quantitative PCR (RT-PCR)

The megagametophytes from *A. angustifolia* seeds were homogenized in liquid nitrogen with pestle and mortar, and total RNA was extracted using the protocol described by Chang et al. (1993). The quantity and purity of the RNA samples were assessed using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, United States); samples with 260/280 nm and 260/230 nm ratios between 1.8–2.2 and 1.6–2.2, respectively, were considered pure enough. The integrity of the samples was confirmed by electrophoresis on a 1.5% (w/v) agarose gel. Total RNA was treated with DNase I (New England Biolabs). Then, first-strand cDNA was synthesized using M-MuLV Reverse Transcriptase (New England Biolabs) and d(t) 20 oligonucleotide, following the manufacturer's instructions.

Gene expression was evaluated through semiquantitative RT-PCR. Nuclease primers were designed using Primer3Plus Program (Untergasser et al., 2007). The endogenous normalization was performed using Ubiquitin 1 gene (Schlögl et al., 2012). The primer sequences are shown in Table 1. The PCR reactions were conducted in a total volume of 25 µL containing 5 µL of 1:10 diluted cDNA, 0.5 U Taq polymerase (Invitrogen), 0.2 mM dNTP, 0.1 µM for a specific sense and anti-sense primers, 5 µL 10× PCR buffer (Invitrogen) and 0.5 mM MgCl₂. The thermal cycle conditions used were: 94°C for 3 min, 94°C for 30 s, 58°C for 30 s, 60°C for 30 s and 72°C

TABLE 1 | Primer sequences used in this work.

Primers	Sequence	Amplified fragment
AaUb1-Fw	5'GTCGGATGTGTTTCATCCTAATG3'	160 bp of the Ubiquitin 1 gene
AaUb1-Rr	5'CTTCTGGATTGCAGGACTTG3'	
Ac520-Fw	5'TAGGGCAATGGTGGTTAATG3'	229 bp of an uncharacterized protein (endonuclease activity)
Ac520-Rv	5'AAATTCTGCTGCCTCATGTGC3'	
Ac114-Fw	5'AGTGCATGAGGCTTACCTTG3'	218 bp of an uncharacterized protein (endonuclease activity)
Ac114-Rv	5'TAACCATTCCCGAACAAGAG3'	
Ac343-Fw	5'GAGATGAAGGTGGAAACACG3'	247 bp of an uncharacterized protein (endonuclease activity)
Ac343-Rv	5'AGACGAATGCTTTCAGTTGC3'	

for 1 min. The numbers of cycles were specific for each pair of primers. The PCR products had a length between 160 and 247 bp. The RT-PCR products were resolved on 1.5% (w/v) agarose gel and stained with ethidium bromide (0.5 µg/mL).

In-gel Nuclease Activity Assays

For *in-gel* nuclease assays, megagametophytes at different stages were ground in liquid nitrogen and homogenized in extraction buffer [10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.1 (w/v) % SDS, 0.1 phenylmethylsulphonyl fluoride (PMSF) from Roche (Mannheim, Germany), and 1 mM dithiothreitol (DTT)]. Equal amounts of protein (15 µg) were incubated for 20 min at 40°C in buffer [0.125 M Tris pH 6.8, 10% (v/v) glycerol, 2% (w/v) SDS, 0.01% (w/v) bromophenol blue] and resolved on 12% SDS-PAGE gels containing 0.3 mg mL⁻¹ herring sperm DNA (Biodynamics, Argentina). For single-stranded DNase activity, DNA was boiled for 5 min immediately prior to pouring the gel. The gels were soaked in 25% 2-propanol and 1 mM EDTA for 15 min to remove SDS as previously reported by Leśniewicz et al. (2010). Subsequently, the gels were incubated overnight in 25 mM sodium acetate-acetic acid buffer [pH 5.5, 0.2 mM DTT and 1% (v/v) Triton X-100] or 10 mM Tris-HCl neutral buffer [pH 8.0, 0.2 mM DTT and 1% (v/v) Triton X-100] at 37°C. *In-gel* assays in the presence of cations were performed as above, in buffer containing 0.1 mM ZnSO₄ or 10 mM CaCl₂. After incubations, the gels were washed for 5 min in cold stop buffer [10 mM Tris-HCl (pH 8.0), 1 mM EDTA]. Nuclease activity was detected as a negatively stained band revealed by staining the gels with 0.01 mg/mL ethidium bromide and photographed using the Box GeneSnap software from Syngene. The band intensity was analyzed using the Gel-Pro Analyzer Software (Media Cybernetics Inc.). All SDS-PAGE results were replicated a minimum of three times.

DNA Isolation and Fragmentation Analysis

Genomic DNA was isolated by the cetyl-trimethyl-ammonium-bromide (CTAB) method (Doyle, 1991). Then, 200 mg of three different megagametophytes were ground with liquid nitrogen into a fine powder and mixed with 400 µL CTAB solution [1.4 M NaCl; 2% (w/v) PVPPM_{40,000}, 20 mM EDTA (pH 8.0), 100 mM

Tris-HCl, pH 8.0; 2% (w/v) CTAB]. The mix was incubated for 15 min at 70°C. An equal volume of chloroform:isoamyl alcohol mixture (24:1) was added and, after shaking gently, the mixture was centrifuged for 10 min at 10,000 g. The upper aqueous phase was removed and the total DNA was precipitated by addition of 700 μ L 70% (v/v) ethanol. DNA was recovered by centrifugation for 2 min at 10,000 g. The yield and quality of the DNA obtained were assessed in a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, United States). For DNA-fragmentation analysis, 20 μ g of each sample was separated on a 2% (w/v) agarose gel and stained with ethidium bromide (0.5 μ g/mL). A Thermo Fisher Scientific™ DNA GeneRuler™ 100 bp was used as a reference.

Protein Concentration

The protein concentrations were determined as described by Bradford (1976), using a Quick Start Bradford Protein Assay Kit 1 (500–0201; Bio-Rad, United States Laboratories) and bovine serum albumin (BSA) as a standard (Bio-Rad Laboratories, United States).

TUNEL Assay on Megagametophyte Cells After Germination

Nuclear DNA fragmentation was detected by TdT-mediated dUTP nick-end labeling (TUNEL) according to the protocol provided by the manufacturer (*In situ* cell detection kit TMR red, Roche, Merck KGaA, Darmstadt, Germany). Briefly, tissue sections from proximal to embryo area were treated with 0.05% Tween 20 in PBS for 15 min at room temperature, to facilitate penetration of the labeling reagents. The slides were incubated in the TUNEL reaction mix at 37°C for 60 min. To prepare negative controls, sections were incubated without the Terminal deoxynucleotide Transferase (TdT) enzyme from the reaction mixture; to obtain positive controls, sections were pretreated with DNase I (image not shown). The percentage of TUNEL positive nuclei, at 0, 14, 28, and 42 DAG, were calculated from 100 nuclei randomly selected, for each section. At least 5 semi thin sections of a different megagametophyte tissue were observed.

Cys-EP Immunological Assays Western Blotting

Megagametophyte tissue of *A. angustifolia* from different DAG and endosperms of *R. communis* from seeds 5 DAG were ground in liquid nitrogen and homogenized in extraction buffer [50 mM Tris-HCl (pH 8), 50 mM DTT, 1 mM EDTA and 1 mM PMSF]. The cellular extracts were centrifuged for 10 min at 14,000 g, 4°C. Western blot analyses were performed as previously described in López-Fernández and Maldonado (2013b), with minor modifications. Equal amounts of protein were separated using SDS-PAGE and electrotransferred to a PVDF membrane (Millipore Corporation, Bedford, MA, United States) at 100 V for 60 min. The membrane was immersed in 3% (w/v) BSA in a TTBS solution [0.2 M Tris-HCl (pH 7.6), 1.37 M NaCl, 0.1% (v/v) Tween-20] overnight at 4°C. The proteins were incubated with a primary antibody raised against purified 35 kDa Cys-EP (Schmid et al., 1998)

diluted 1:1000 in 1% (w/v) BSA in TTBS for 2 h at room temperature and subjected to five 5 min rinses in a TTBS solution. The membrane was then incubated with a secondary alkaline phosphatase-conjugated goat anti-rabbit antibody (Sigma A3587, Merck KGaA, Darmstadt, Germany) diluted 1:5000 in TTBS for 1:30 h at room temperature. The secondary antibody was detected with NBT/BCIP (Promega, Madison, WI, United States).

In situ Immunolocalization

Immunolocalization was carried out according to the protocol described by Schmid et al. (1999) and López-Fernández and Maldonado (2013b). Briefly, after blocking with 1% (w/v) BSA in PBS for 90 min, the slides were incubated with anti-Cys-EP (dilution 1:100 in 0.1% BSA/PBS) overnight at 4°C. After washing in PBS plus 0.05% (v/v) Tween 20 (PBST) three times for 10 min, the slides were incubated with a fluorescent anti-rabbit ALEXA 488 IgG (Invitrogen, Thermo Fisher Scientific, Waltham, MA, United States) antibody applied 1:1000 in 0.1% BSA/PBS for 1 h at room temperature in the dark. After additional rinses in PBST, sections were examined by epifluorescence and light microscopy.

Microscope Settings

Images were obtained by epifluorescence and light microscopy with an Axioskop 2 microscope (Carl Zeiss, Jena, Germany). All images were captured with an EOS 1000D camera (Canon, Tokyo, Japan), analyzed using the AxioVision 4.8.2 software package (Carl Zeiss, Jena, Germany), and compiled (Photoshop version CS6; Adobe Systems, San Jose, CA, United States). Rhodamine filters (excitation 520–560 nm, emission 570–620 nm) and DAPI filters (excitation 340–390 nm, emission 420–470 nm) were used to examine samples by TUNEL assay, whereas Alexa filters (excitation 450–490 nm, emission 515–565 nm) were used to examine samples for immunofluorescence assays.

RESULTS

Histochemical Staining and Analysis of Tissue Sections Revealed Progressive Megagametophyte Degradation, Mobilization of Reserves and Cells Undergoing Programmed Cell Death (PCD) During Germination

During germination, the megagametophyte changed its color progressively from white and bright to brownish. **Figure 1A** shows the progressive cell death observed in the megagametophyte, from the innermost layers to the outer ones, using Evans Blue dye: living cells are able to exclude the dye (and thus cells remain unstained), whereas dead cells lose membrane integrity and stain blue. At 0 DAG, only the remnants of the cell layers close to the embryo appear stained, and no extensive cell death is observed until well after the mobilization of reserves has finalized. At 42 DAG, cell death has been initiated in the entire tissue.

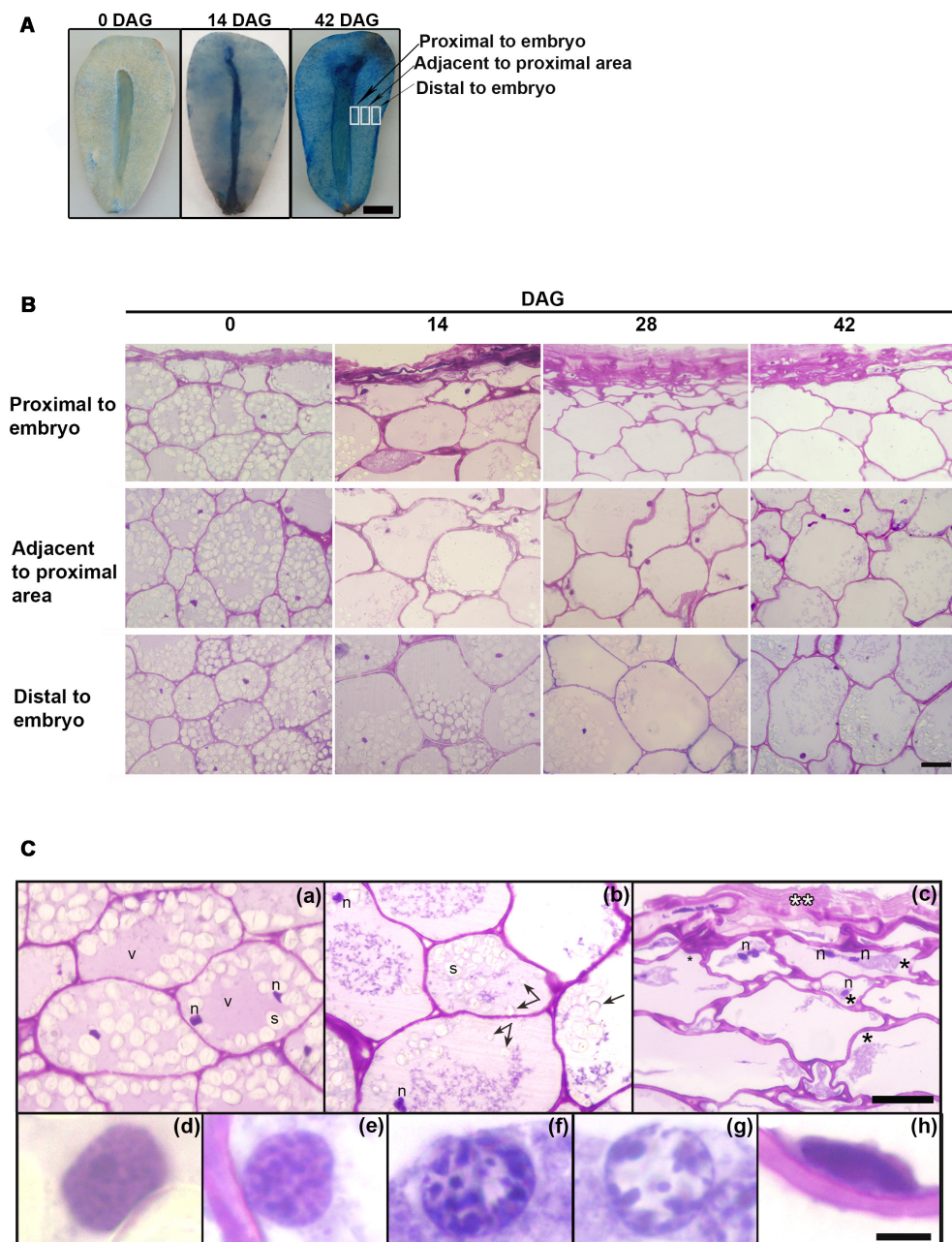


FIGURE 1 | (A) Evans Blue staining at 0 and 42 DAG in the megagametophyte after extraction of the embryo. In each case the figure is a representative result of observation of at least 6 seeds. Scale bar = 0.5 cm. **(B)** Changes occurring from the proximal to distal zones in the megagametophyte, at 0, 14, 28, and 42 DAG. Scale bar: 50 μ m. **(C)** In detail, megagametophyte tissue (14 DAG). **(a)**, Distal section of the tissue; **(b,c)**, Adjacent to proximal sections; in c collapsed cell area. **(Ca,c)**, some cells show two nuclei. **(d–h)**, Progressive changes in nuclear morphology. The black asterisk indicates plasmolysis, the double white asterisk indicates crushed cells, and the black arrow indicates amyloplast discharged into central vacuole. Scale bar: **(a–c)** = 50 μ m, **(d–h)** = 5 μ m. Abbreviations: n, nucleus; v, vacuole; s, starch. Each image is a representative result of observation of at least 30 semi-thin sections of megagametophyte tissue at different DAG.

Tissue death began in the proximal sector of the embryo and extended distally (**Figures 1B,C**). At 0 DAG, the number of fully collapsed cell layers was small. As the degradation of the tissue progressed, the cell walls weakened and lost rigidity, and finally the cells collapsed; the remnants of the degraded cell walls persisted in the crushed cell layers proximal to the embryo

(**Figure 1B**). Vacuolar proteins were stained with Amido black at 0 DAG. Storage proteins were relatively scarce and diluted in the water of the central vacuole. Once the germination started, the vacuolar proteins were completely consumed (**Supplementary Figure S1**). During germination, nuclei showed alterations in size and morphology (**Figures 1B,C**): initially large and round, nuclei

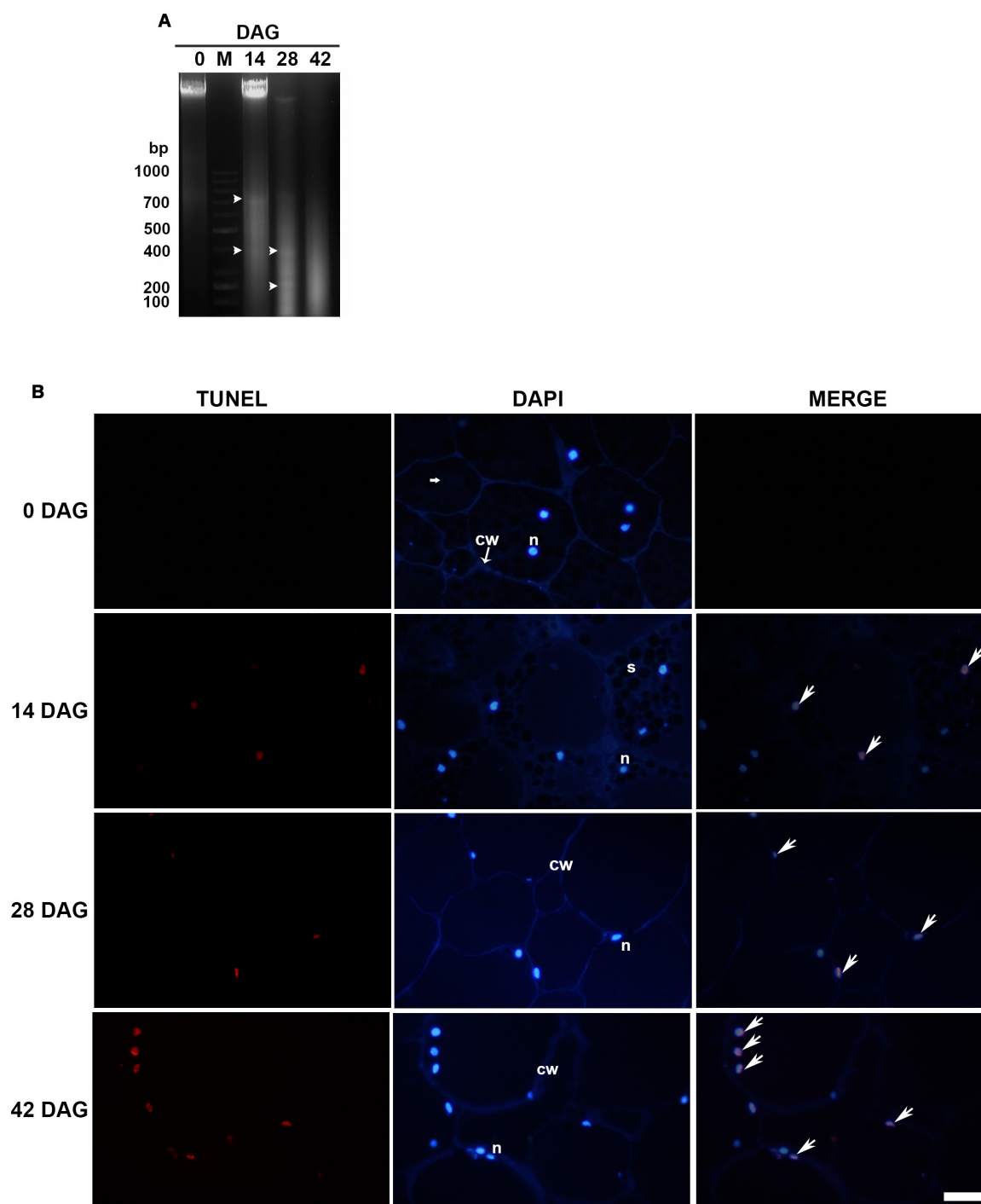


FIGURE 2 | (A) Genomic DNA fragmentation assay. Analyses were repeated at least three times on independent biological samples, and representative results are shown. Lane M. 1 kb ladder, Lanes 1, 3, 4, and 5 correspond to 0, 14, 28, and 42 DAG, respectively. Bands are indicated by arrowheads **(B)** TUNEL assay (left column) and DAPI staining (central column) were performed on LRW tissue sections at 0, 14, 28, and 42 DAG. Merged images (right column) confirmed that DAPI co-labeled TUNEL-positive nuclei. Scale bar = 50 μ m. Each image is a representative result of observation of at least 5 semi-thin sections of megagametophyte tissue at different DAG.

reduced in size and became fusiform. Also, a progressive increase in chromatin condensation was observed. The reserves were progressively consumed; specifically, starch inside amyloplasts

was degraded. Starch depletion began in the proximal sector to the embryo prior to 14 DAG approximately, advancing in the distal direction (**Figure 1B** and **Supplementary Figure S2**). After

hydrolysis of the starch, amyloplasts were discharged into the central vacuole (Figure 1Cb and Supplementary Figure S2). Disorganization of the cytoplasm occurred later with the collapse of the central vacuole, the general degradation of cytoplasmic organelles and plasmolysis (Figures 1Cb,c).

The nuclear dismantling was associated with cytoplasmic events during plant PCD. However, chromatin was partially adhered to the nuclear membrane until very advanced the processes of mobilization of reserves (Figures 1Cd-h).

During Storage Mobilization, DNA Fragmentation Accompanied the Progressive Cellular Changes Observed in the Megagametophyte

Analysis of genomic DNA integrity of the total tissue by electrophoresis on agarose gel between 14 and 42 DAG revealed three bands of approximately 700, 400, and 200 bp, respectively. The 700 bp band was clearly weaker between 28 and 42 DAG. Fragmentation was not significant before 14 DAG. Also a faint DNA smearing was observed at 14, 28, and 42 DAG and one well-bound band of high-molecular weight corresponding to intact DNA was visible at 0 and 14 DAG (Figure 2A). The lack of detection of a clear laddering could be interpreted as the result of the DNA analysis of a tissue with areas that shown different timing of PCD. To evaluate the *in situ* detection of DNA damage, a TUNEL assay was performed. At 14 DAG, the first TUNEL-positive signals were detected at the innermost layers of the proximal area and later advanced toward the distal area, continuously increasing the number of affected nuclei (Figure 2B). Percentage of nuclei labeled was 33, 56, and 92% at 14, 28, and 42 DAG, respectively. Analysis of DAPI-stained nuclei by fluorescence microscopy exhibited the progressive changes in the nuclear morphology, as above-mentioned.

In the Cells Undergoing PCD, Cys-EP Accumulated in the Cytoplasm and Cell Walls

The protein extracts from 0, 14, 28, and 42 DAG were separated by SDS-PAGE and electrophoretically transferred to a PVDF membrane. The western blot clearly showed the presence of the immature and mature forms (approximately 45 and 38 kDa, respectively) of Cys-EP (Figure 3A). A major band of 38 kDa, which corresponds to the active form of CysEP, was observed at 14 and 28 DAG. At 42 DAG, neither of the two bands appeared revealed.

The *in situ* accumulation pattern of Cys-EP at 14 DAG was studied on longitudinal sections of the megagametophyte from the proximal and adjacent sectors of the embryo to the proximal cell layers. As mentioned above, the cell walls progressively lost rigidity (Figure 1B). Figure 3B indicates that, in cells proximal to the embryo, Cys-EP immunolocalized in ricinosomes mixed with vacuolar content and highly vesiculated cytoplasm (a); in vacuolated cells (i.e., with vacuole not collapsed), Cys-EP localized in the parietal cytoplasm, next to the cell wall (b).

In the Megagametophyte, Zn^{2+} Induced the Activity of Nucleases During Germination

Nuclease activity was determined using *in-gel* activity assay with single-stranded (ssDNA) or double-stranded DNA (dsDNA) as substrate at acidic (pH 5) or neutral (pH 8) conditions (Figure 4). When ssDNA was used as substrate, the activities of three nucleases with molecular masses of 65, 43, and 35 kDa respectively were detected. The activities of n43 and n35 were strongly enhanced by Zn^{2+} at 28 and 42 DAG. It is worth noting that, although n43 digested ssDNA, in the presence of both Ca^{2+} and Zn^{2+} ions, its ability to digest dsDNA was stimulated only by Zn^{2+} ions at acidic conditions. No bands were obtained when the *in-gel* activity assay was performed at neutral conditions with dsDNA as substrate. The enhancement of Zn^{2+} nuclease activity occurred simultaneously with nuclear DNA fragmentation, i.e., when the band of high molecular weight had disappeared and two bands of low molecular weight (400 and 200 bp, respectively) were markedly visible (Figure 2A).

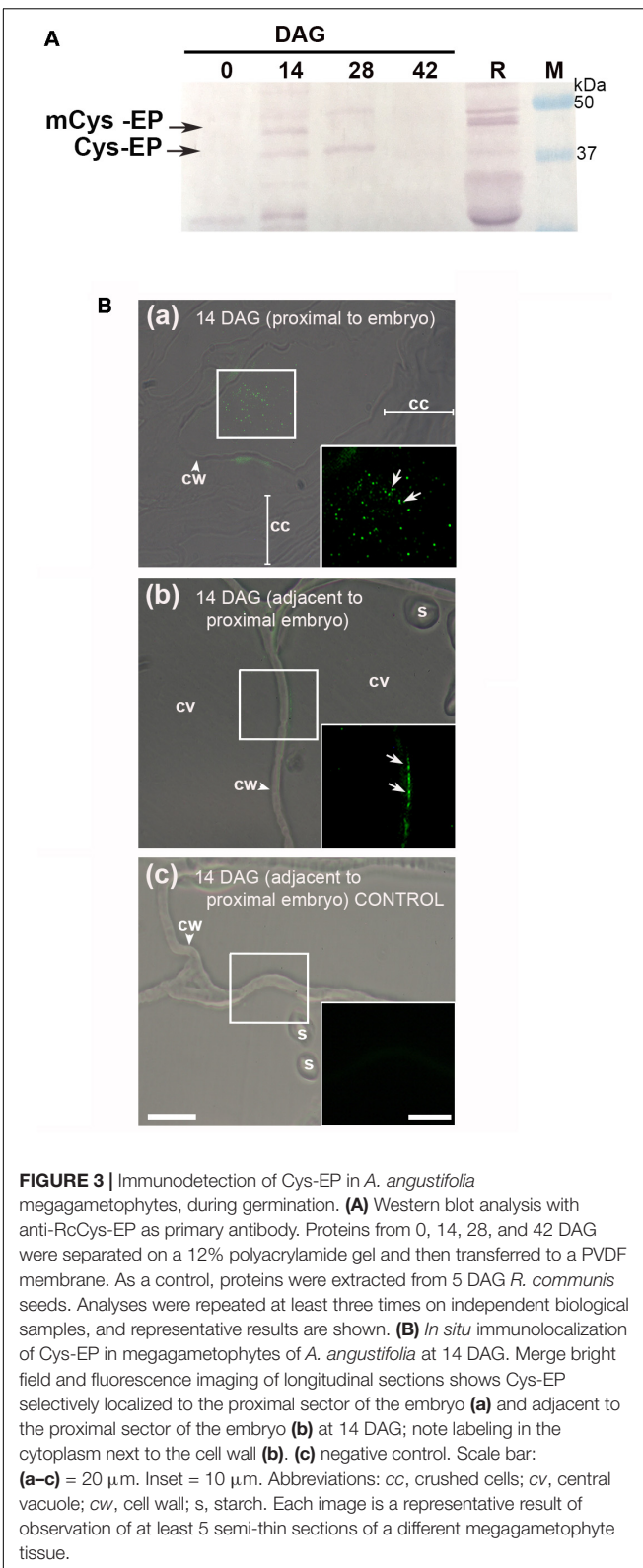
Correlated with the increase in DNA-activity, there was a reduction in both DNA and RNA contents, especially at 14 DAG (Figure 4B). During germination, changes were detected in the RNA and soluble protein contents, which, at 42 DAG, reached values close to zero. In addition, the DNA content decreased drastically, reaching values close to 5 $\mu\text{g/g}$ at 42 DAG (i.e., 2.6 times when compared to the control) (Figure 2A).

In silico Analysis of Putative Nucleases in *Araucaria angustifolia*

Putative nuclease-encoding genes from *A. angustifolia* were searched within different nucleotide sequence databases. Although no nuclease-encoding genes from *A. angustifolia* were identified, three sequences of transcribed RNAs encoding for putative nucleases from *Araucaria cunninghamii*, a closely related species, were found. These three sequences were obtained from the European Nucleotide Archive¹ and were submitted under Accession Numbers GCKF01036520.1, GCKF01039343.1, and GCKF01021114.1, respectively. All these sequences were identified from a leaf by transcriptomic analysis, which indicates that they come from genes that are expressed at least in leaf tissue.

A sequence analysis of these three putative nuclease-encoding genes was performed using the BLAST tool available at the National Center for Biotechnology Information (NCBI) website (NCBI Resource Coordinators 2016). Firstly, the analysis of the amino acid sequence of the predicted protein encoded by the transcribed RNA of Acc. N° GCKF01036520.1 (herein after referred to as putative protein Ac520) revealed high similarity to several Tudor Staphylococcal nucleases (Tudor-SN). The sequence with the highest similarity to Ac520 was a Tudor-SN from *Wollemia nobilis* (Uniprot Acc. N° A0A0C9RQ39), sharing 98.4% of identical amino acids. Ac520 also showed 81% identity to a Tudor-SN from *Picea abies* (Uniprot Acc. No. Q0JRI3), and 64.5, 64.4, and 63.5% identity to ribonuclease TUDOR1

¹<https://www.ebi.ac.uk/ena>



(Uniprot Acc. No. Q8VZG7), isoform 2 of ribonuclease TUDOR 1 (Uniprot Acc. No. Q8VZG7-2) and ribonuclease TUDOR 2 (Uniprot Acc. No. Q9FLT0) from *Arabidopsis thaliana*,

respectively. Furthermore, Ac520 has the conserved domains of Tudor-SN nucleases (**Figure 5Ai**), that is, four SN domains at the N-terminus and a Tudor domain at the C-terminus (Liu et al., 2010). Tudor-SN proteins have been shown to be involved in the control of seed germination in *A. thaliana* (Liu et al., 2010) and in the regulation of PCD in plants (Sundström et al., 2009; Coll et al., 2010; Reape and McCabe, 2010; Tsiatsiani et al., 2011). The analysis of the amino acid sequence of the predicted protein encoded by the transcribed RNA of Acc. No. GCKF01039343.1 (herein after referred to as putative protein Ac343) showed the presence of a S1-P1 nuclease domain (**Figure 5Aii**). Ac343 showed 63.1 and 60.8% identity to ENDO4 (Uniprot Acc. No. F4JLL0) and ENDO2 (Uniprot Acc. No. Q9C9G4) from *A. thaliana*, respectively. This last nuclease was involved in RNA, ssDNA, and dsDNA degradation, with a preference for ssDNA and RNA (Ko et al., 2012). Finally, the analysis of the amino acid sequence of the predicted protein encoded by the transcribed RNA of Acc. No. GCKF01021114.1 (herein after referred to as putative protein Ac114) revealed the presence of a conserved DNase-RNase domain (**Figure 5Aiii**). This domain is characteristic of a family of bifunctional nucleases having both DNase and RNase activity. In fact, Ac114 shares 80% of amino acid identity with a predicted bifunctional nuclease from *Picea sitchensis* (Uniprot Acc. No. A9NUL3) previously reported by Ralph et al. (2008). Ac114 also showed 67.6% identity to BBD2 (Uniprot Acc. No. Q93VH2) and 63.7% identity to BBD1 (Uniprot Acc. No. Q9FWS6) from *A. thaliana*.

Expression Levels of Putative Nuclease Genes During Germination in Megagametophytes of *Araucaria angustifolia*

To evaluate the expression of genes from *A. angustifolia* encoding for orthologous nucleases of Ac520, Ac343, and Ac114, primers were designed to carry out semi-quantitative RT-PCR. It should be noted that *A. angustifolia* and *A. cunninghamii* are very closely related species and therefore the orthologous sequences are not expected to have significant differences. The transcript levels of the genes encoding Ac520, Ac343, and Ac114 were analyzed in the megagametophyte of *A. angustifolia* at 0, 14, and 28 years 42 DAG (**Figure 5B**). Expression of the Ac520 and Ac114 genes increased along the germination process, with higher expression levels at 28 and 42 DAG, whereas that of the Ac343 gene was only detected at 28 and 42 DAG.

DISCUSSION

Nuclease activation, DNA fragmentation and reserve mobilization in the megagametophyte of *Araucaria angustifolia* occurred simultaneously during the first 4 weeks following germination. During storage mobilization, DNA fragmentation accompanied the progressive cellular changes observed in the cells of the megagametophyte. On the basis of these results, we propose that the pathway of cell death in the *A. angustifolia* megagametophyte is PCD. As mentioned, to date,

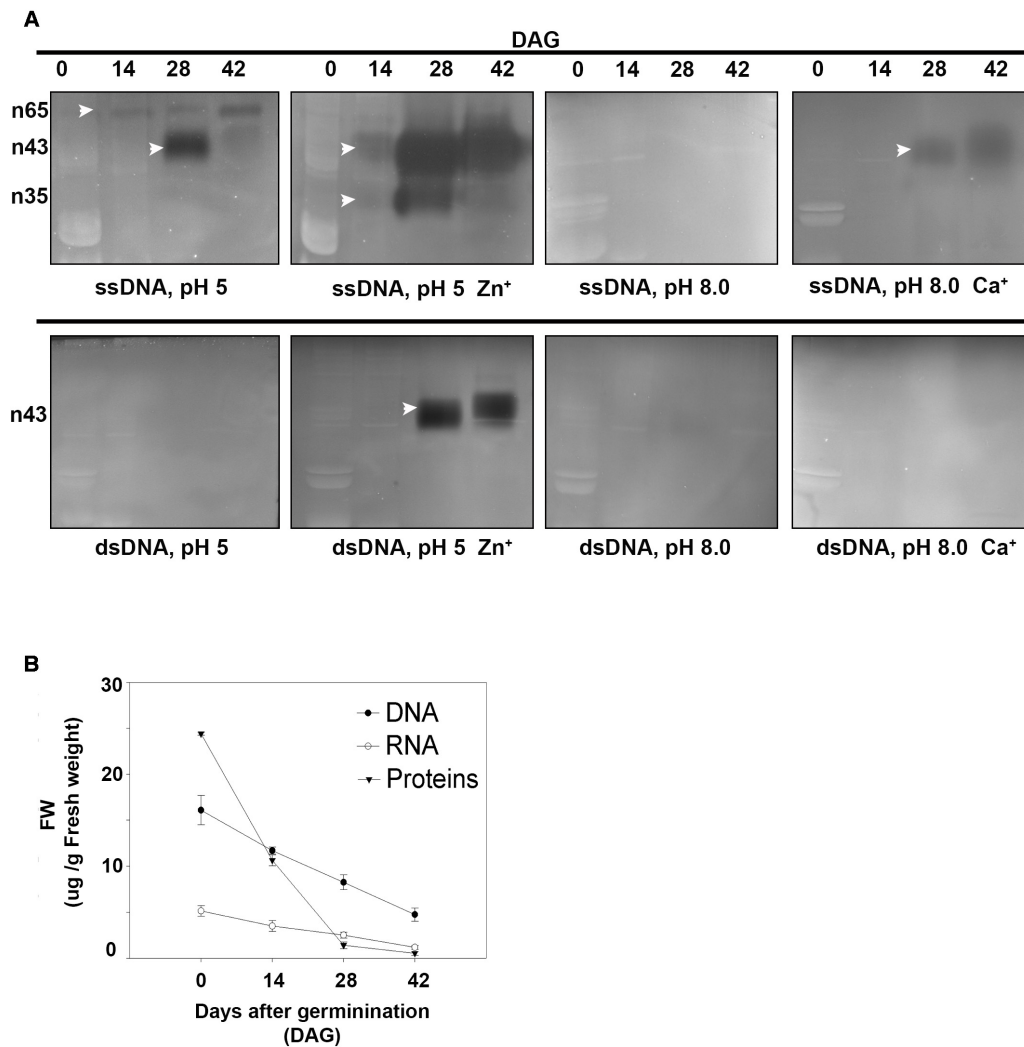


FIGURE 4 | (A) Determination of nuclease activity in *A. angustifolia* megagametophytes at 0, 14, 28, and 42 days after germination (DAG) by using the *in-gel* activity assay with ssDNA or dsDNA as substrate. See section “Materials and Methods” for detailed reaction conditions. Analyses were repeated at least three times on independent biological samples, and representative results are shown. **(B)** Changes in soluble protein, DNA and RNA in the megagametophyte, during germination. The three parameters exhibited a marked decrease initiated before 14 DAG. DNA, RNA, and soluble proteins were extracted at 0, 14, 28, and 42 DAG, quantified, and normalized on fresh weight. Data are represented as the mean \pm standard error from three independent biological replicates.

the dismantling of the megagametophyte during germination has been studied in only two species of Gymnosperms: *Araucaria bidwillii* and *Picea glauca*. In *A. bidwillii*, a species that is very closely related to *A. angustifolia* and also produces starchy seeds, cell death has been reported to occur through necrosis and probably also PCD in some cells (Casani et al., 2009). In *Picea glauca*, the megagametophyte is a tissue whose cells store proteins and lipids and lacks vacuoles, and the reserve mobilization pattern and the PCD pathway are similar to those described in the aleurone layer of cereals or in the endosperm of tomato (He and Kermode, 2003).

Autophagy is a process known to mediate the degradation of residual proteins and aggregates of insoluble proteins and lipids, and to remove damaged organelles (Levine and Klionsky, 2004; Mizushima, 2007; Rodriguez-Navarro and Cuervo, 2010).

In addition, autophagic assimilation and reprocessing can maintain cellular homeostasis, responding to environmental changes, but can also function in association with the PCD process (Kourtis and Tavernarakis, 2009; Kuma and Mizushima, 2010). Although it is known that autophagy also mediates bulk degradation of the cytosol and organelles in plants, its role in plastid catabolism is largely unknown (Wada et al., 2008). In the megagametophyte of *A. angustifolia*, we observed that, after starch hydrolysis, autophagy was responsible for the final degradation of amyloplasts. This process seems to be similar to that occurring in chloroplasts of *Arabidopsis* leaves during senescence (Wada et al., 2008), although this issue needs further investigation.

KDEL-tailed Cys-EPs digest extensin, thus supporting the final cell collapse during PCD. Cys-EPs were detected for the

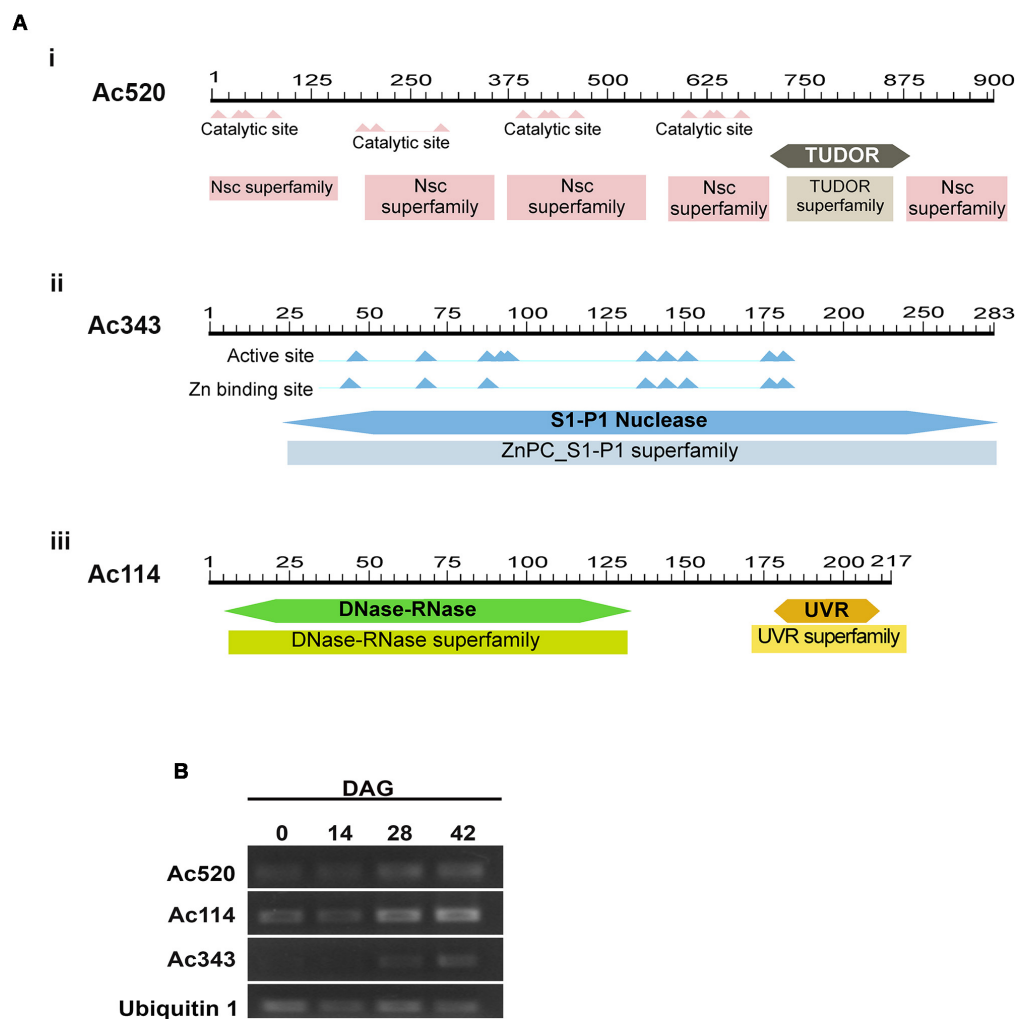


FIGURE 5 | (A) *In silico* analysis of putative nuclease in *Araucaria angustifolia*. **(i)** Representation of the Ac520 protein showing the Staphylococcal nuclease domains at the N terminus and a Tudor domain toward the C terminus. **(ii)** Representation of the Ac343 protein showing the S1-P1 Nuclease domain. **(iii)** Representation of the Ac114 protein showing the DNase-RNase domain. **(B)** Expression levels of the putative nuclease genes during the germination process in *A. angustifolia*. Amplified products of the Ac520, Ac114, and Ac343 genes by semiquantitative RT-PCR. As a control for constitutive expression, a fragment of the Ubiquitin 1 gene was simultaneously amplified. The results are representative of at least three independent biological replicates.

first time in the endosperm of *Ricinus communis* (Schmid et al., 2001) and the epigeal cotyledons of *Vigna mungo* (Toyooka et al., 2001) during germination. Cys-EPs have also been identified in the megagametophyte of *Picea glauca* (He and Kermode, 2003). Here, we detected a Cys-EP in the megagametophyte of *A. angustifolia* during germination by using an antibody raised against a Cys-EP purified from ricinosomes of the endosperm of *Ricinus communis* seeds during germination. This Cys-EP was immunolocalized in ricinosomes mixed with vacuolar content and in the parietal cytoplasm, next to the cell wall. By western blot, we recognized both the proform and mature form of the enzyme at 14 and 28 DAG but not at 42 DAG. According to Schmid et al. (1998), the other bands that we detected correspond to a precursor protease, a C-terminally truncated active form and to degradation products.

In the present study, two Zn^{2+} -dependent nucleases of 35 and 43 kDa were induced at 28–42 DAG at acid pH. Nucleases of the S1/P1 family are thought to be similar to nucleases type I, showing maximal activity at acidic pH and Zn^{2+} dependency (Sugiyama et al., 2000).

Here, we found three sequences of transcribed RNAs, Ac520, Ac343, and Ac114, encoding for putative nucleases. Ac520 revealed high similarity to several Tudor-SN, with highest similarity to a Tudor-SN from *Wollemia nobilis* (Uniprot Acc. No. A0A0C9RQ39), with which it shared 98.4% of identical amino acids. *Wollemia nobilis* is the only species of the genus *Wollemia* that also belongs to the family Araucariaceae (Division Pinophyta-Order Pinales) (Jones et al., 1995). According to Setoguchi et al. (1998), the Araucariaceae are well defined by the *rbcl* sequence, and their monophyly is supported by a bootstrap value of 100%.

Likewise, Ac520 showed 81% identity to a Tudor-SN from *Picea abies* (Uniprot Acc. No. Q0JRI3), a species of the family Pinaceae, which also belongs to the Division Pinophyta-Order Pinales; this species is phylogenetically and temporally very distant from *Araucaria*. In fact, the Pinaceae diverged from the lineage ultimately leading to *Araucaria* in the Late Carboniferous to Early Permian periods, approximately 300–250 million years ago (Gernandt et al., 2008; Leslie et al., 2012). Ac520 also exhibited identity to TUDOR ribonucleases (Uniprot Acc. No. Q8VZG7, Uniprot Acc. No. Q8VZG7-2, Uniprot Acc. No. Q9FLT0) from *Arabidopsis thaliana*.

It is important to note that *A. thaliana* is an Angiosperm species phylogenetically very distant from *Araucaria*. Furthermore, Ac520 has the conserved domains of Tudor-SN, that is, four SN domains at the N-terminus and a Tudor domain at the C-terminus (Liu et al., 2010). Tudor-SN proteins have been shown to be involved in the control of seed germination in *A. thaliana* (Liu et al., 2010) and in the regulation of PCD in plants (Sundström et al., 2009; Coll et al., 2010; Reape and McCabe, 2010; Tsiatsiani et al., 2011). Ac343 showed the presence of a S1-P1 nuclease domain, and 63.1 and 60.8% identity to ENDO 4 (Uniprot Acc. No. F4JJL0) and ENDO2 (Uniprot Acc. No. Q9C9G4) from *Arabidopsis thaliana*, respectively. This last nuclease is involved in RNA, ssDNA, and dsDNA degradation, with a preference for ssDNA and RNA (Ko et al., 2012). Ac343 also presented 55% amino acid identity to the best characterized plant S1-like nucleases, ZEN1 of *Zinnia elegans* and *Arabidopsis* ENDO1 (also named bifunctional nuclease1, BFN1). Previous reports have demonstrated that BFN1 and ZEN1 are involved in different forms of PCD (Pérez-Amador et al., 2000; Ko et al., 2012; Lesniewicz et al., 2013). Finally, Ac114 revealed the presence of a conserved DNase-RNase domain, which is characteristic of a family of bifunctional nucleases having both DNase and RNase activity. In fact, Ac114 shared 80% of amino acid identity with a predicted bifunctional nuclease from *Picea sitchensis*, a species of the family Pinaceae (Uniprot Acc. No. A9NUL3) previously reported by Ralph et al. (2008). Ac114 also showed 67.6% identity to BBD2 (Uniprot Acc. No. Q93VH2) and 63.7% identity to BBD1 (Uniprot Acc. No. Q9FWS6) from *A. thaliana*.

The results of the present study add knowledge on the dismantling of the megagametophyte of mature starchy seeds in species of the genus *Araucaria*, a storage tissue that stores

starch as the main reserve substance, as well as on the PCD pathway, by revealing new insights into the role of nucleases and the expression patterns of putative nuclease genes during germination.

AUTHOR CONTRIBUTIONS

SM and ML-F conceived, designed and coordinated the project, and initiated the project. ML-F coordinated the field work and sampling. LM, MC, FR, LF, and ML-F performed laboratory work. LM, MC, FR, LF, SM, and ML-F performed the data analysis. SM and ML-F wrote the first draft of the paper. All authors contributed to discussing the results and editing the paper.

FUNDING

This work was supported by Universidad de Buenos Aires (UBACYT 20020100100232 to SM) and Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET. Res. 7879/14 IA to ML-F and 810/13.P IP 0465 to SM), Argentina.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.01275/full#supplementary-material>

FIGURE S1 | Amido black staining at 0, 14, and 42 DAG in the megagametophyte sections. Amido black stains total proteins. They were visible at 0 DAG. At 14 and 42 DAG, vacuoles shown that storage proteins are consumed early. In all sections, nuclear proteins were also dyed. Abbreviations: cc, crushed cells; v, vacuole; cw, cell wall; am, amyloplasts. Scale bar = 50 μ m. Each image is a representative result of observation of at least 30 semi-thin sections of a different megagametophyte tissue at different DAG.

FIGURE S2 | Lugol staining at 0, 14, and 42 DAG in the megagametophyte sections. Starch was progressively consumed during germination. Starch was histochemically identified by lugol staining. Vacuolar transport of entire amyloplast can be observed (arrows). Amyloplasts finished being dismantled and starch totally consumed within the central vacuole. Abbreviations: cc, crushed cells; v, vacuole; cw, cell wall; am, amyloplasts. Scale bar = 50 μ m. Each image is a representative result of observation of at least 30 semi-thin sections of a different megagametophyte tissue at different DAG.

REFERENCES

- Bethke, P., Lonsdale, J., Fath, A., and Jones, R. (1999). Hormonally regulated programmed cell death in barley aleurone cells. *Plant Cell* 11, 1033–1046. doi: 10.1105/tpc.11.6.1033
- Bewley, J. D., Bradford, K. J., Hilhorst, H. W. M., and Nonogaki, H. (2013). *Seeds: Physiology of Development, Germination and Dormancy*, 3rd Edn. Berlin: Springer Science & Business Media, doi: 10.1007/978-1-4614-4693-4
- Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem* 72, 248–254. doi: 10.1016/0003-2697(76)90527-3
- Burrieza, H. P., López-Fernández, M. P., and Maldonado, S. (2014). Analogous reserve distribution and tissue characteristics in quinoa and grass seeds suggest convergent evolution. *Front. Plant Sci.* 5:546. doi: 10.3389/fpls.2014.00546
- Cardemil, L., and Reiner, A. (1982). Changes of *Araucaria araucana* seed reserves during germination and early seedling growth. *Can. J. Bot.* 60, 1629–1638. doi: 10.1139/b82-211
- Cardemil, L., and Varner, J. E. (1984). Starch degradation metabolism towards sucrose synthesis in germinating *Araucaria araucana* seeds. *Plant Physiol.* 76, 1047–1054. doi: 10.1104/pp.76.4.1047
- Casani, S., Fontanini, D., Capocchi, A., Lombardi, L., and Gallechi, L. (2009). Investigation on cell death in the megagametophyte of *Araucaria bidwillii* Hook. post-germinated seeds. *Plant Physiol. Biochem.* 47, 599–607. doi: 10.1016/j.plaphy.2009.02.012

- Chang, S., Puryear, J., and Cairney, J. (1993). A simple and efficient method for isolating RNA from pine trees. *Plant Mol. Biol. Rep.* 11, 113–116. doi: 10.1007/BF02670468
- Coll, N. S., Vercammen, D., Smidler, A., Clover, C., Van Breusegem, F., Dangel, J. L., et al. (2010). Arabidopsis type I metacaspases control cell death. *Science* 330, 1393–1397. doi: 10.1126/science.1194980
- Domínguez, F., and Cejudo, F. J. (2014). Programmed cell death (PCD): an essential process of cereal seed development and germination. *Front. Plant Sci.* 5:366. doi: 10.3389/fpls.2014.00366
- Doyle, J. (1991). "DNA Protocols for Plants," in *Molecular Techniques in Taxonomy*, M. H. Godfrey, A. W. B. Johnston, J. P. W. Young (Berlin: Springer), 283–293. doi: 10.1007/978-3-642-83962-7_18
- Fath, A., Bethke, P., Lonsdale, J., Meza-Romero, R., and Jones, R. (2000). Programmed cell death in cereal aleurone. *Plant Mol. Biol.* 44, 255–266. doi: 10.1023/A:1026584207243
- Gernandt, D. S., Magallón, S., Geadá López, G., Zerón Flores, O., Willyard, A., and Liston, A. (2008). Use of simultaneous analyses to guide fossil-based calibrations of pinaceae phylogeny. *Int. J. Plant Sci.* 169, 1086–1099. doi: 10.1086/590472
- Gietl, C., and Schmid, M. (2001). Ricinosomes: an organelle for developmentally regulated programmed cell death in senescing plant tissues. *Naturwissenschaften* 88, 49–58. doi: 10.1007/s001140000203
- Harris, K. F., Esbroeck, Z. P., and Duffus, J. E. (1995). Moderate-temperature polymerization of LR white in a nitrogen atmosphere. *Microsc. Res. Tech.* 32, 264–265. doi: 10.1002/jemt.1070320312
- He, X., and Kermode, A. R. (2003). Proteases associated with programmed cell death of megagametophyte cells after germination of white spruce (*Picea glauca*) seeds. *Plant Mol. Biol.* 52, 729–744. doi: 10.1023/A:1025008117046
- He, X., and Kermode, A. R. (2010). Programmed cell death of the megagametophyte during post-germinative growth of white spruce (*Picea glauca*) seeds is regulated by reactive oxygen species and the ubiquitin-mediated proteolytic system. *Plant Cell Physiol.* 51, 1707–1720. doi: 10.1093/pcp/pcq130
- Jensen, W. A. (1962). *Botanical Histochemistry: Principles and Practice*. San Francisco, CA: W.H. Freeman & Co.
- Jones, W., Hill, K., and Allen, J. (1995). *Wollemia nobilis*, a new living Australian genus and species in the Araucariaceae. *Telopea* 6, 173–176. doi: 10.7751/telopea19953014
- Ko, C. Y., Lai, Y. L., Liu, W. Y., Lin, C. H., Chen, Y. T., Chen, L. F. O., et al. (2012). Arabidopsis ENDO2: its catalytic role and requirement of n-glycosylation for function. *J. Agric. Food Chem.* 60, 5169–5179. doi: 10.1021/jf300945c
- Kourtis, N., and Tavernarakis, N. (2009). Autophagy and cell death in model organisms. *Cell Death. Differ.* 16, 21–30. doi: 10.1038/cdd.2008.120
- Kuma, A., and Mizushima, N. (2010). Physiological role of autophagy as an intracellular recycling system: with an emphasis on nutrient metabolism. *Semin. Cell Dev. Biol.* 21, 683–690. doi: 10.1016/j.semcdb.2010.03.002
- Lambert, R., Quiles, F. A., Cabello-Díaz, J. M., and Piedras, P. (2014). Purification and identification of a nuclease activity in embryo axes from French bean. *Plant Sci.* 224, 137–143. doi: 10.1016/j.plantsci.2014.04.017
- Leslie, A. B., Beaulieu, J. M., Rai, H. S., Crane, P. R., Donoghue, M. J., and Mathews, S. (2012). Hemisphere-scale differences in conifer evolutionary dynamics. *Proc. Natl. Acad. Sci. U.S.A.* 109, 16217–16221. doi: 10.1073/pnas.1213621109
- Lesniewicz, K., Karlowski, W. M., Pienkowska, J. R., Krzykowski, P., and Poreba, E. (2013). The plant S1-Like nuclease family has evolved a highly diverse range of catalytic capabilities. *Plant Cell Physiol.* 54, 1064–1078. doi: 10.1093/pcp/pct061
- Leśniewicz, K., Pienkowska, J., and Poręba, E. (2010). Characterization of nucleases involved in seedling development of cauliflower. *J. Plant Physiol.* 167, 1093–1100. doi: 10.1016/j.jplph.2010.03.011
- Levine, B., and Klionsky, D. J. (2004). Development by self-digestion: molecular mechanisms and biological functions of autophagy. *Dev. Cell* 6, 463–477. doi: 10.1016/S1534-5807(04)00099-1
- Liu, S., Jia, J., Gao, Y., Zhang, B., and Han, Y. (2010). The AtTudor2, a protein with SN-Tudor domains, is involved in control of seed germination in Arabidopsis. *Planta* 232, 197–207. doi: 10.1007/s00425-010-1167-0
- López-Fernández, M. P., and Maldonado, S. (2013a). Programmed cell death during quinoa perisperm development. *J. Exp. Bot.* 64, 3313–3325. doi: 10.1093/jxb/ert170
- López-Fernández, M. P., and Maldonado, S. (2013b). Ricinosomes provide an early indicator of suspensor and endosperm cells destined to die during late seed development in quinoa (*Chenopodium quinoa*). *Ann. Bot.* 112, 1253–1262. doi: 10.1093/aob/mct184
- Mella, R. A., Sánchez, R. A., and Maldonado, S. (1995). Phytochrome-induced structural changes and protein degradation prior to radicle protrusion in *Datura ferox* seeds. *Can. J. Bot.* 73, 1371–1378. doi: 10.1139/b95-149
- Mizushima, N. (2007). Autophagy: process and function. *Genes Dev.* 21, 2861–2873. doi: 10.1101/gad.1599207
- Owens, J. N., Morris, S. J., and Misra, S. (1993). The ultrastructural, histochemical, and biochemical development of the post-fertilization megagametophyte and the zygotic embryo of *Pseudotsuga menziesii*. *Can. J. For. Res.* 23, 816–827. doi: 10.1139/x93-106
- Panza, V., Lainez, V., Maroder, H., Prego, I., and Maldonado, S. (2002). Storage reserves and cellular water in mature seeds of *Araucaria angustifolia*. *Bot. J. Linn. Soc.* 140, 273–281. doi: 10.1046/j.1095-8339.2002.00093.x
- Pérez-Amador, M., Abler, M. L., De Rocher, E. J., Thompson, D. M., van Hoof, A., LeBrasseur, N. D., et al. (2000). Identification of BFN1, a bifunctional nuclease induced during leaf and stem senescence in Arabidopsis. *Plant Physiol.* 122, 169–180. doi: 10.1104/pp.122.1.169
- Ralph, S. G., Chun, H. J. E., Kolosova, N., Cooper, D., Oddy, C., Ritland, C. E., et al. (2008). A conifer genomics resource of 200,000 spruce (*Picea* spp.) ESTs and 6,464 high-quality, sequence-finished full-length cDNAs for *Sitka spruce* (*Picea sitchensis*). *BMC Genomics* 9:484. doi: 10.1186/1471-2164-9-484
- Reape, T. J., and McCabe, P. F. (2010). Apoptotic-like regulation of programmed cell death in plants. *Apoptosis* 15, 249–256. doi: 10.1007/s10495-009-0447-2
- Rodríguez-Navarro, J. A., and Cuervo, A. M. (2010). Autophagy and lipids: tightening the knot. *Semin. Immunopathol.* 32, 343–353. doi: 10.1007/s00281-010-0219-7
- Sabelli, P. A. (2012). Replicate and die for your own good: endoreduplication and cell death in the cereal endosperm. *J. Cereal Sci.* 56, 9–20. doi: 10.1016/j.jcs.2011.09.006
- Sakamoto, W., and Takami, T. (2014). Nucleases in higher plants and their possible involvement in DNA degradation during leaf senescence. *J. Exp. Bot.* 65, 3835–3843. doi: 10.1093/jxb/eru091
- Schlögl, P. S., dos Santos, A. L. W., do Nascimento Vieira, L., Floh, E. I. S., and Guerra, M. P. (2012). Gene expression during early somatic embryogenesis in Brazilian pine (*Araucaria angustifolia* (Bert) O. Ktze). *Plant Cell. Tissue Organ Cult.* 108, 173–180. doi: 10.1007/s11240-011-0023-7
- Schmid, M., Simpson, D., and Gietl, C. (1999). Programmed cell death in castor bean endosperm is associated with the accumulation and release of a cysteine endopeptidase from ricinosomes. *Proc. Natl. Acad. Sci. U.S.A.* 96, 14159–14164. doi: 10.1073/pnas.96.24.14159
- Schmid, M., Simpson, D., Kalousek, F., and Gietl, C. (1998). A cysteine endopeptidase with a C-terminal KDEL motif isolated from castor bean endosperm is a marker enzyme for the ricinosomes, a putative lytic compartment. *Planta* 206, 466–475. doi: 10.1007/s004250050423
- Schmid, M., Simpson, D. J., Sarioglu, H., Lottspeich, F., and Gietl, C. (2001). The ricinosomes of senescing plant tissue bud from the endoplasmic reticulum. *Proc. Natl. Acad. Sci. U.S.A.* 98, 5353–5358. doi: 10.1073/pnas.061038298
- Setoguchi, H., Asakawa Osawa, T., Pintaud, J. C., Jaffré, T., and Veillon, J. M. (1998). Phylogenetic relationships within Araucariaceae based on rbcL gene sequences. *Am. J. Bot.* 85, 1507–1516. doi: 10.2307/2446478
- Sugiyama, M., Ito, J., Aoyagi, S., and Fukuda, H. (2000). Endonucleases. *Plant Mol. Biol.* 44, 387–397. doi: 10.1023/A:1026504911786
- Sundström, J. F., Vaculova, A., Smertenko, A. P., Savenkov, E. I., Golovko, A., Minina, E., et al. (2009). Tudor staphylococcal nuclease is an evolutionarily conserved component of the programmed cell death degradome. *Nat. Cell Biol.* 11, 1347–1354. doi: 10.1038/ncb1979
- Tompsett, P. B. (1984). Desiccation studies in relation to the storage of Araucaria seed. *Ann. Appl. Biol.* 105, 581–586. doi: 10.1111/j.1744-7348.1984.tb03085.x
- Toyooka, K., Okamoto, T., and Minamikawa, T. (2001). Cotyledon cells of *Vigna mungo* seedlings use at least two distinct autophagic machineries for degradation of starch granules and cellular components. *J. Cell Biol.* 154, 973–982. doi: 10.1083/jcb.200105096
- Tsiatsiani, L., Van Breusegem, F., Gallois, P., Zavialov, A., Lam, E., and Bozhkov, P. V. (2011). Metacaspases. *Cell Death Differ.* 18, 1279–1288. doi: 10.1038/cdd.2011.66

- Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., and Leunissen, J. A. M. (2007). Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* 35, W71–W74. doi: 10.1093/nar/gkm306
- Wada, S., Ishida, H., Izumi, M., Yoshimoto, K., Ohsumi, Y., Mae, T., et al. (2008). Autophagy plays a role in chloroplast degradation during senescence in individually darkened leaves. *Plant Physiol.* 149, 885–893. doi: 10.1104/pp.108.130013
- Young, T. E., and Gallie, D. R. (1999). Analysis of programmed cell death in wheat endosperm reveals differences in endosperm development between cereals. *Plant Mol. Biol.* 39, 915–926. doi: 10.1023/A:1006134027834
- Young, T. E., and Gallie, D. R. (2000a). Programmed cell death during endosperm development. *Plant Mol. Biol.* 44, 283–301. doi: 10.1023/A:1026588408152
- Young, T. E., and Gallie, D. R. (2000b). Regulation of programmed cell death in maize endosperm by abscisic acid. *Plant Mol. Biol.* 42, 397–414.
- Zheng, Y., Zhang, H., Deng, X., Liu, J., and Chen, H. (2017). The relationship between vacuolation and initiation of PCD in rice (*Oryza sativa*) aleurone cells. *Sci. Rep.* 7:41245. doi: 10.1038/srep41245

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 © 2018 Moyano, Correa, Favre, Rodríguez, Maldonado and López-Fernández. 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.



Evidence for the Extensive Conservation of Mechanisms of Ovule Integument Development Since the Most Recent Common Ancestor of Living Angiosperms

Gontran Arnault¹, Aurélie C. M. Vialette¹, Amélie Andres-Robin¹, Bruno Fogliani², Gildas Gâteblé² and Charles P. Scutt^{1*}

¹ Laboratoire Reproduction et Développement des Plantes, École Normale Supérieure de Lyon, Université Claude Bernard Lyon 1, CNRS, INRA, Université de Lyon, Lyon, France, ² Équipe ARBOREAL, "Agriculture Biodiversité et Valorisation", Institut Agronomique Néo-Calédonien (IAC), Païta, New Caledonia

OPEN ACCESS

Edited by:

Annette Becker,
Justus-Liebig-Universität Gießen,
Germany

Reviewed by:

Charles Gasser,
University of California, Davis,
United States
Masaru Ohme-Takagi,
Saitama University, Japan

*Correspondence:

Charles P. Scutt
charlie.scutt@ens-lyon.fr

Specialty section:

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

Received: 23 June 2018

Accepted: 28 August 2018

Published: 19 September 2018

Citation:

Arnault G, Vialette ACM,
Andres-Robin A, Fogliani B,
Gâteblé G and Scutt CP (2018)
Evidence for the Extensive
Conservation of Mechanisms of Ovule
Integument Development Since
the Most Recent Common Ancestor
of Living Angiosperms.
Front. Plant Sci. 9:1352.
doi: 10.3389/fpls.2018.01352

The ovules and seeds of most angiosperm groups are enclosed by two integuments, whose evolutionary origins are considerably separated in time, as the inner integument arose over 300 million years ago (MYA) in an ancestor of all living seed plants, while the outer integument arose, perhaps as recently as 164 MYA, in an ancestor of all living angiosperms. Studies of the model angiosperm *Arabidopsis thaliana* indicate that the mechanisms of development of the inner and outer integuments depend on largely different sets of molecular players. However, it was not known, in most cases, whether these differences were already present in early flowering plants, or arose later in the *Arabidopsis* lineage. Here, we analyze the expression patterns of integument regulators in *Amborella trichopoda*, the likely sister to all other living angiosperms. The data obtained indicate that regulators of the YABBY, KANADI, and homeodomain-leucine zipper class III transcription factor families have largely conserved their integument-specific expression profiles in the *Amborella* and *Arabidopsis* lineages since the most recent common ancestor (MRCA) of living angiosperms. We identified only one case, involving the paralogous genes *ETTIN* and *AUXIN RESPONSE FACTOR4*, in which integument-specific expression patterns had clearly diverged between *Amborella* and *Arabidopsis*. We use the data obtained to partially reconstruct molecular mechanisms of integument development in the MRCA of living angiosperms and discuss our findings in the context of alternative hypotheses for the origin of the angiosperm outer integument.

Keywords: integument, *Amborella trichopoda*, ovule, angiosperms, HD-ZIP III, YABBY, KANADI, AUXIN RESPONSE FACTOR

INTRODUCTION

The ovules and seeds of most seed plants are covered by one or two integuments. These maternal tissues function to (1) protect the internal tissues of the ovule and later the seed, (2) define a route, via the micropyle, for pollen or pollen-tube entry, and (3) contribute in many cases to the regulation of seed hydration and dormancy (Linkies et al., 2010). Gymnosperms possess a single

integument, whereas the majority of angiosperm groups, including the most basally diverging of these, possess two integuments, of which the inner integument is considered homologous to the single integument of gymnosperms. The origin of the inner integument therefore dates from before the separation of the living angiosperm and gymnosperm lineages, believed to have occurred over 300 million years ago (MYA), while the outer integument must have arisen somewhere along the angiosperm stem lineage, perhaps as recently as ~164 MYA: a reasonable estimated date for the most recent common ancestor (MRCA) of living angiosperms (Moore et al., 2010). Within angiosperms, several groups have undergone a secondary reduction to a single integument (Endress, 2011), notably including the asterids, which contain over 80,000 species.

Much of what is known of the molecular mechanisms of integument development comes from the study of the model angiosperm *Arabidopsis thaliana* (hereafter referred to as *Arabidopsis*), which possesses anatropous, bitegmic ovules. Ovule initiation in *Arabidopsis* begins with the expression of the homeobox transcription factor *WUSCHEL* (*WUS*) (see Mathews and Kramer, 2012), which is known to promote cell proliferation in meristematic tissues. The ovule thus shares characteristics of meristems, such as the stem apical meristem and floral meristem, both of which produce lateral organs on their flanks. In the case of the bitegmic ovule, the lateral organs produced in this way are the inner and outer integuments, which arise from the chalazal tissue in the central region of the elongating ovule primordium. The inner integument arises first, followed closely by the outer integument (Endress, 2011).

Early expression of *WUS* at the apex of the *Arabidopsis* ovule primordium defines the presumptive nucellar tissue. The outgrowth of the integuments from the chalazal region, immediately below the nucellus, depends on the expression of three homeodomain-leucine zipper class III (HD-ZIP III) transcription factors, *PHABULOSA* (*PHB*), *PHAVOLUTA* (*PHV*), and *CORONA* (*CNA*) (Kelley et al., 2009), which act redundantly, and together with other factors, to limit the basal spread of *WUS* expression (Yamada et al., 2016). Expression of these HD-ZIP III factors defines the adaxial (towards the growth axis) zone of the presumptive inner integument. The inner and outer integument primordia are associated with the formation of auxin maxima, and it has been proposed that *ETTIN/AUXIN RESPONSE FACTOR3* (*ETT/ARF3*) and *ABERRANT TESTA SHAPE/KANADI4* (*ATS/KAN4*), both of which are expressed in the abaxial (away from the growth axis) domain of the inner integument, act to remove auxin from the zone between the two integument primordia (Kelley et al., 2012). Interestingly, Lora et al. (2015) found that in certain species of *Prunus* whose ovules contain only one integument, expression of the *ETT* ortholog was absent from the chalaza and inner integument. These authors accordingly proposed the loss of *ETT* expression to represent a potential evolutionary mechanism for the reduction from two integuments to one in these species.

HD-ZIP III and KANADI family factors are believed to control abaxial–adaxial tissue polarity in lateral organs through

the regulation of a common set of direct target genes, many of which are related to auxin signaling or dynamics (Reinhart et al., 2013; Huang et al., 2014). HD-ZIP III factors regulate these targets positively to promote adaxial tissue identity, while KANADI factors regulate the same genes negatively to promote abaxial tissue identity. In *Arabidopsis*, there is a clear division of labor between the inner and outer integuments, as *ATS* is the principal KANADI family member expressed in the inner integument (McAbee et al., 2006), while this role is played redundantly by *KAN1* and *KAN2* in the outer integument (Eshed et al., 2004; McAbee et al., 2006). In the HD-ZIP III family, *PHB*, *PHV*, and *CNA* are expressed specifically in the inner integument, while *REV* is expressed in both integuments (Sieber et al., 2004b; Kelley et al., 2009).

Tissue outgrowth in angiosperm lateral organs, including for example leaves and carpels, is typically associated with the expression of YABBY transcription factors. These factors are expressed in the abaxial tissue domain, and can act either positively or negatively on the transcription of downstream targets, apparently through protein–protein interactions with distinct sets of co-factors (Bonaccorso et al., 2012; Simon et al., 2017). Despite their abaxial-specific expression profiles, YABBYs are believed not to define abaxial tissue identity *per se*, but rather to facilitate lateral organ outgrowth by promoting the expression of *WUS*-related transcription factors in marginal meristems (Nakata et al., 2012). The YABBY gene *INNER NO OUTER* (*INO*), is expressed specifically in the abaxial zone of the *Arabidopsis* outer integument, and an orthologue of this gene shows a largely similar expression pattern in the basally diverging angiosperm *Nymphaea* (Nymphaeales, Nymphaeaceae) (Yamada et al., 2003, 2011), suggesting *INO*'s tissue-specific role to have been conserved from early stages of angiosperm evolution.

With very few exceptions (such as that of *INO*, mentioned above), the mechanisms of integument development have been investigated to date uniquely in *Arabidopsis*, and it is therefore not known, in most cases, which of these were already present in early flowering plants and which evolved later in the *Arabidopsis* lineage. In the present work, we investigate the conservation of mechanisms of integument development from very early stages of angiosperm evolution by analyzing the expression patterns of the orthologs of *Arabidopsis* integument regulators in the basally diverging angiosperm *Amborella trichopoda* (hereafter referred to as *Amborella*). Most recent molecular phylogenetic studies place *Amborella*, the only known representative of Amborellales, as sister to all other living angiosperms (Stevens, 2001 onwards). According to these studies, Amborellales is the first of the three earliest diverging “ANA-grade” angiosperm orders, the remaining two being Nymphaeales and Austrobaileyales. *Amborella* is a dioecious scrambling shrub, endemic to the subtropical rainforests of the southern pacific island of New Caledonia. Its female flowers produce 5–6 unfused carpels, each containing a single, pendant, bitegmic ovule. These ovules are of more-or-less orthotropous symmetry, though show a distinct longitudinal curvature, which may be due to developmental constraints within the ovary, or may represent the residual effect of

an ancestral anatrochous symmetry (Endress and Igersheim, 2000).

Here, we focus on integument regulators of the YABBY, KANADI, HD-ZIP III, and ARF transcription factor families. We use gene expression patterns to identify molecular mechanisms of integument development that have probably been conserved since the MRCA of living flowering plants, and others that appear to have undergone changes in either the *Amborella* or *Arabidopsis* lineages. We discuss the data obtained in the context of alternative hypotheses for the origin of the outer integument in a distant common ancestor of living angiosperms.

MATERIALS AND METHODS

Phylogenetic Reconstruction

Phylogenetic reconstructions in the KANADI and HD-ZIP III families were performed using a wide taxonomic sampling of angiosperms. Sequences for inclusion in phylogenetic analyses were selected by TBLASTN searching (Altschul et al., 1997) of the NCBI/NIH non-redundant nucleotide database¹ and the *Amborella* genome database² using *Arabidopsis* integument development proteins. Protein sequences were aligned using MUSCLE in the SeaView phylogeny package (Gouy et al., 2010). Sites were selected using G-BLOCKS, using all three options provided to minimize the stringency of selection. Phylogenies were generated from amino acid alignments in PhyML (Guindon et al., 2009) using the following parameters: model: LG; branch support: aLRT; amino acid equilibrium frequencies: model-given; invariable sites: optimized; across-site rate variation: optimized; tree searching operations: best of NNI and SPR; and starting tree: BioNJ.

Gene Expression Studies

Full-length nucleotide sequences corresponding to *Amborella* orthologs or pro-orthologs of *Arabidopsis* integument regulators were PCR-amplified from an *Amborella* female flower cDNA library (Fourquin et al., 2005) using the primers shown in **Supplementary Table S1**. Digoxigenin-labeled riboprobes were generated from these and used in non-radioisotopic *in situ* hybridization to tissue sections of *Amborella* flower buds at approximately Stage 7 (Buzgo et al., 2004), and of female flowers at anthesis, using the protocol given by Vialette-Guiraud et al. (2011). Gene expression patterns were observed and photographed under bright field illumination using a Leica Axio Imager M2 inverted microscope fitted with a Leica AxioCam MRc digital camera.

Anatomical Observations

Amborella ovule anatomy was revealed in sections of fixed female flowers, prepared and photographed as for *in situ* hybridization. These sections were stained with 0.05% (w/v) toluidine Blue-0 in 0.1 M sodium phosphate buffer (pH 6.8).

RESULTS

Clear Orthologs of Several *Arabidopsis* Integument Regulators Can Be Identified in *Amborella*

Several previous studies have focused on the molecular phylogeny of gene families involved in the regulation of integument development. Accordingly, the *Amborella* ortholog of the *Arabidopsis* YABBY-family integument regulator *INO* has previously been reported (Finet et al., 2016), as have those of the paralogous and partially redundant auxin response family regulators *ETT* and *AUXIN RESPONSE FACTOR4 (ARF4)* (Finet et al., 2010). However, published phylogenies of the KANADI and HD-ZIP III families have not to our knowledge previously included sequences from *Amborella*.

Phylogenetic reconstructions of the HD-ZIP III family in the present study (**Supplementary Figure S1**), generated from the protein alignment shown in **Supplementary Figure S2**, succeeded in identifying clear *Amborella* orthologs of the *Arabidopsis* genes *REV* and *CNA*, and a clear pro-ortholog of the *Arabidopsis* genes *PHB* and *PHV*, which accordingly appear to be derived from a duplication within the crown group of living angiosperms. In all these cases, the *Amborella* genes identified as orthologs or pro-orthologs of *Arabidopsis* integument regulators occupied basal positions within their respective clades, in agreement with the likely phylogenetic placement of *Amborella*.

Phylogenetic reconstruction of the KANADI family (**Supplementary Figure S3**), generated from the protein alignment shown in **Supplementary Figure S4**, succeeded in identifying a clear *Amborella* ortholog of the *Arabidopsis* inner integument regulator *ATS (KAN4)*, which grouped in a small basal clade of *ATS*-like sequences. Relationships of gene orthology between *Amborella* sequences and the *Arabidopsis* outer integument regulators *KAN1* and *KAN2* were less clear in phylogenetic reconstructions.

The results obtained in these reconstructions, combined with those of previously published work, permitted detailed analyses in the present study of the expression in *Amborella* female flower tissues of the genes: *Atr.INO* (XM_006840616), *Atr.ATS*, (XM_006856161.3), *Atr.PHB/PHV* (XM_006845952.3), *Atr.REV* (XM_006846714.3), *Atr.ARF4* (XM_020671056), and *Atr.ETT* (XM_011627988.2). The longer alternative splicing variant of *Atr.ARF4* (Finet et al., 2010) was used for *in situ* hybridization studies.

The *Amborella* Orthologue of the YABBY Gene *INO* Is Narrowly Expressed in the Abaxial Domain of the Outer Integument, As Is *INO* in *Arabidopsis*

Atr.INO proved to be specifically expressed in the most abaxial cell layer of the outer integument. Its expression was first noted in the outer integument, towards the chalazal-end of the ovary (**Figure 1A**), spreading towards the micropyle-end with the growth of the outer integument (**Figure 1B**). This expression

¹ <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

² <http://genomevolution.org/CoGe/>

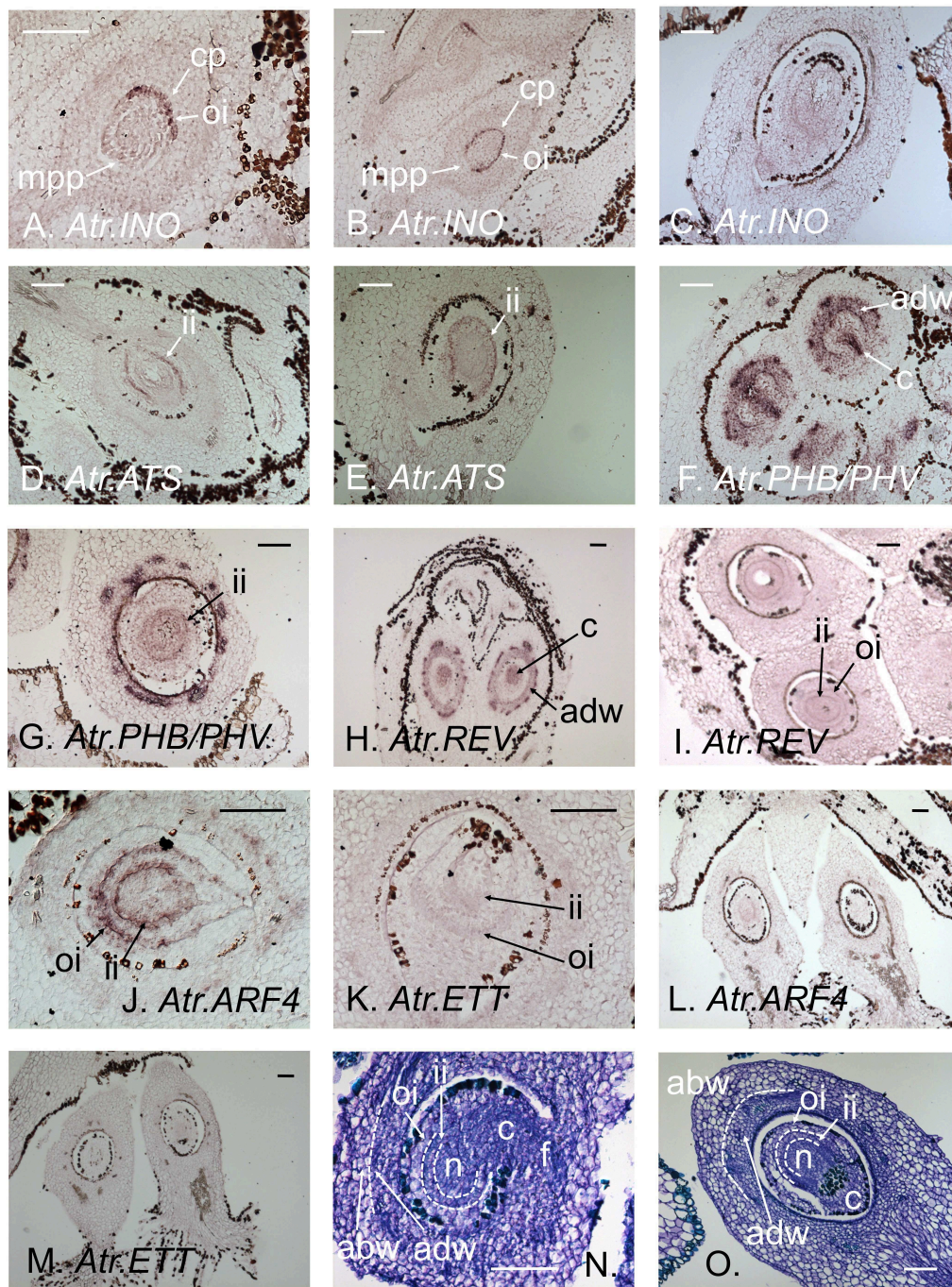


FIGURE 1 | *In situ* hybridizations of integument regulators to female *Amborella* flower sections. **(A)** *Atr.INO* expression in the growing outer integument at early Stage 7. **(B)** *Atr.INO* expression in the near fully grown outer integument at late Stage 7. **(C)** *Atr.INO* expression is no longer apparent in the outer integument (LS) in mature flowers. **(D)** *Atr.ATS* expression in the inner integument at floral bud Stage 7. **(E)** *Atr.ATS* expression in the inner integument persists in the mature flower. **(F)** Narrow *Atr.PHB/PHV* expression in the chalaza and strong expression in the adaxial domain of the ovary wall at early Stage 7. **(G)** Weak *Atr.PHB/PHV* expression persists in the abaxial domain of the inner integument at late Stage 7. **(H)** *Atr.REV* is extensively expressed in the chalaza, and strongly expressed in the adaxial domain of the ovary wall at early Stage 7. **(I)** Weak *Atr.REV* expression persists in the adaxial domains of both integuments at late Stage 7. **(J)** *Atr.ARF4* expression in the abaxial domain of both integuments at early Stage 7. **(K)** Very weak *Atr.ETT* expression in the abaxial domain of both integuments at early Stage 7. **(L)** *Atr.ARF4* expression is no longer apparent in the integuments in the mature flower. **(M)** *Atr.ETT* expression is no longer apparent in the integuments in the mature flower. **(N,O)** Toluidine-blue stained sections of Stage 7 and mature ovules, respectively, showing anatomical details. Tissues zones in panels **(N,O)** have been delimited with dotted white lines for clarity. Developmental stages are as defined by Buzgo et al. (2004). abw, abaxial domain of the ovary wall; adw, adaxial domain of the ovary wall; c, chalaza; cp, chalazal pole; f, funiculus; ii, inner integument; mpp, micropyle pole; n, nucellus; and oi, outer integument. Scale bars = 100 μ m.

pattern closely resembles that of *INO* in *Arabidopsis*. *Atr.INO* expression was stronger in early stages of *Amborella* ovule development, becoming absent in the mature ovule (Figure 1C). These data reinforce observations in *Nymphaea* (Yamada et al., 2003) that already suggested *Arabidopsis INO* to have conserved its function in outer integument development from early stages of angiosperm evolution. Expression patterns of *INO* orthologs in *Arabidopsis* and *Amborella* ovules, together with those of the other factors studied (below), are summarized in Figure 2.

The *Amborella* Orthologue of the KANADI Family Gene *ATS* (*KAN4*) Is Expressed Abaxially in the Inner Integument, As Is *ATS* in *Arabidopsis*

Atr.ATS was found to be expressed specifically in the abaxial domain of the inner integument (Figure 1D), as is *ATS* in *Arabidopsis*. This expression persisted into the final stages of ovule development in almost mature flowers (Figure 1E). No expression of *Atr.ATS* was apparent in any female reproductive

tissue other than the inner integument. These data strongly suggest *ATS* orthologs to have conserved their function in inner integument development in both the *Amborella* and *Arabidopsis* lineages since the MRCA of living flowering plants.

Amborella HD-ZIP III Genes Show Integument-Specific Profiles Similar to Those of Their *Arabidopsis* Orthologs

We found *Atr.PHB/PHV*, the *Amborella* pro-ortholog of *Arabidopsis PHB* and *PHV*, to be strongly expressed at early stages of development in the chalazal region of the ovule, which gives rise to the integuments, and very strongly expressed in the adaxial tissues of the ovary wall (Figure 1F). At later developmental stages, *Atr.PHB/PHV* appeared weakly expressed in the adaxial zone of the inner integument (Figure 1G). At early stages of ovule development, *Atr.REV* showed a similar expression pattern to *Atr.PHB/PHV* in the chalaza and ovary wall, though arguably with a slightly wider expression in the chalaza (Figure 1H). At later stages of ovule development, reduced levels of *Atr.REV*

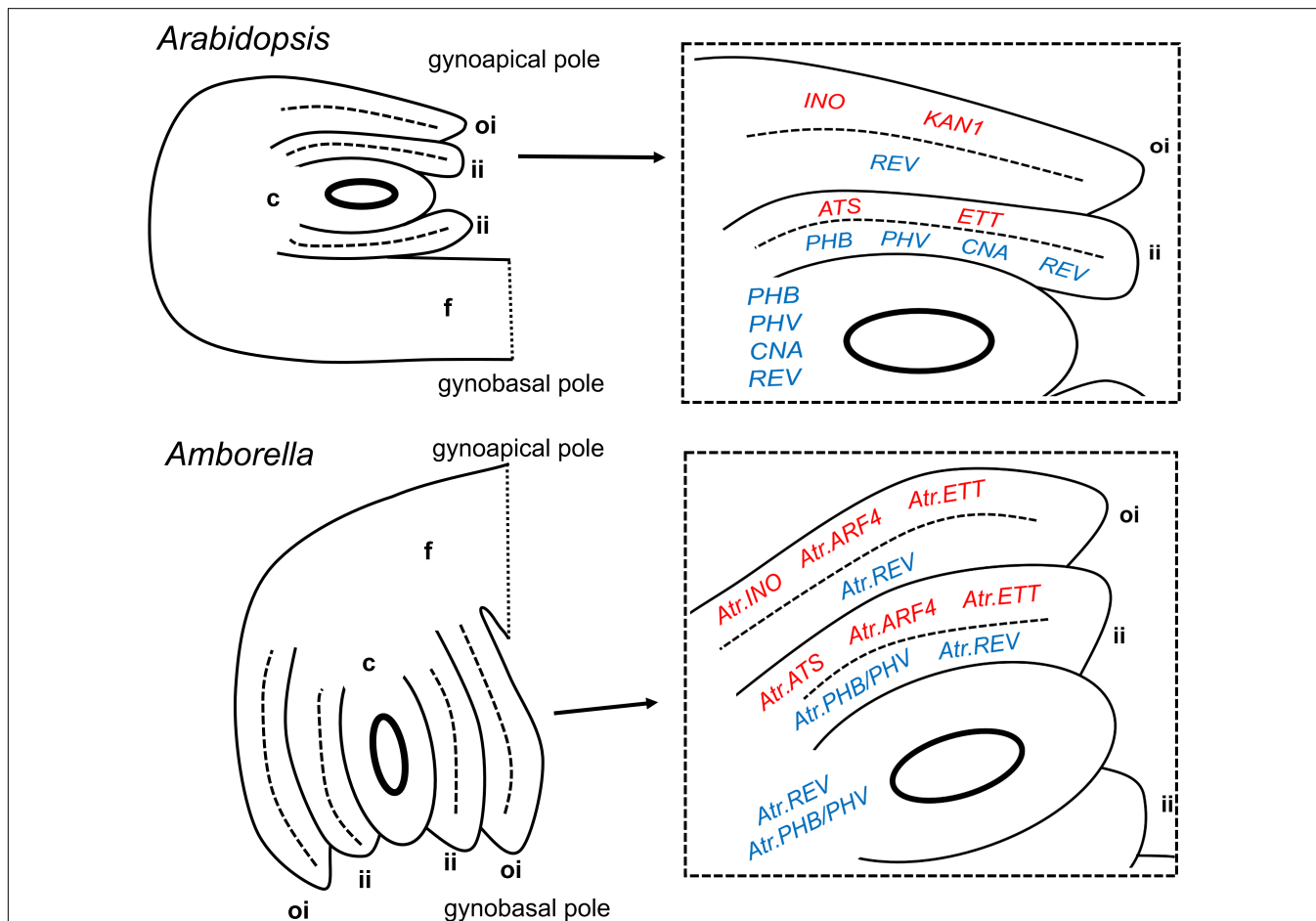


FIGURE 2 | Summary of expression patterns of ovule integument regulators in *Arabidopsis thaliana* and *Amborella trichopoda*. Abaxial-promoting genes are shown in red in enlarged diagrams, adaxial-promoting genes are shown in blue. Mature ovules are shown, though the genes indicated may be maximally expressed in the tissues shown at earlier stages. Entire ovules (on the left) are shown in their natural orientations relative to the longitudinal axis of the gynoecium/carpel. c, chalaza; f, funiculus; ii, inner integument; and oi = outer integument.

expression were visible in the adaxial regions of both integuments (**Figure 1I**). The association of *Atr.PHB/PHV* expression with the inner integument, and that of *Atr.REV* with both integuments (**Figures 1G,I**), suggests the paralog-specific conservation of function between the *Amborella* and *Arabidopsis* lineages since their separation at the base of the living flowering plant clade. *Atr.PHB/PHV* and *Atr.REV* expression patterns are nonetheless very similar, and these genes are clearly expressed together in parts of the chalaza, the inner integument and the abaxial region of the ovary wall.

The *Amborella* Orthologs of the Paralogous Auxin Response Factors *ETT* and *ARF4* Are Expressed Abaxially in Both Integuments, Showing Partial Conservation With Expression Patterns in *Arabidopsis*

ETT and *ARF4* act redundantly in the definition of abaxial tissue identity in the *Arabidopsis* leaf (Pekker et al., 2005), though *ETT* appears to play the major role in this process in both the carpel wall and inner integument. An earlier study of the orthologs of these factors in *Amborella* (Finet et al., 2010) suggested that *Atr.ARF4* may show a higher level of expression than *Atr.ETT* in female reproductive tissues and perhaps therefore play the major developmental role. In the present study, both *Atr.ARF4* and *Atr.ETT* were found to be expressed at early stages in the abaxial domain of both integuments (**Figures 1J,K**), with *Atr.ARF4* giving the stronger hybridization signal. Expression levels of both genes diminished in mature integuments (**Figures 1L,M**). A switch may therefore have occurred in the major paralog active in the inner integument, with *ETT* playing the principal role in *Arabidopsis*, while *Atr.ARF4* is more highly expressed in *Amborella*. In addition, whereas the role of *ETT* is limited

to the inner integument in *Arabidopsis*, *Atr.ARF4* and *Atr.ETT* are expressed in the abaxial domains of both integuments in *Amborella*.

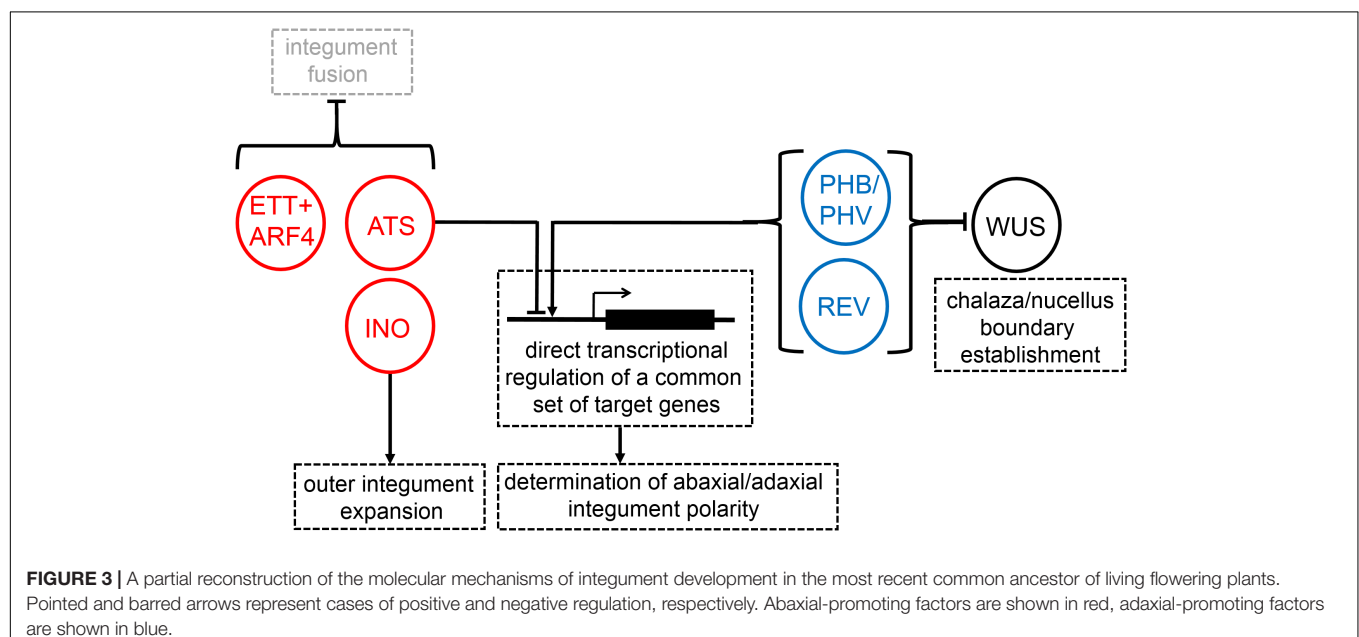
It is interesting to note that *Arabidopsis* *ETT* has been found to interact physically with *ATS*, both of which are expressed specifically in the inner integument (Kelley et al., 2012). *Atr.ETT* and *Atr.ARF4* may similarly interact with *Atr.ATS* in the *Amborella* inner integument, but this interaction could not occur in the outer integument as *Atr.ATS* is not apparently expressed in that tissue (**Figures 1D,E**). Of course, *Atr.ETT* and *Atr.ARF4* might interact in the outer integument with other *Amborella* KANADI proteins (**Supplementary Figure S4**) whose expression patterns were not investigated in the present study.

Toluidine-blue stained sections of the carpel wall and ovule in a Stage 7 (**Figure 1N**) and a mature (**Figure 1O**) female flower are provided to help interpret anatomical details of *in situ* hybridization images shown in **Figures 1A–M**.

DISCUSSION

Reconstructed Gene Expression Patterns Suggest That Largely Distinct Sets of Regulators Controlled Inner and Outer Integument Development in Early Flowering Plants, As In Present-Day *Arabidopsis*

The data presented here strongly suggest the extensive conservation of integument-specific developmental roles of genes of the YABBY, KANADI, and HD-ZIP III families in both the *Amborella* and *Arabidopsis* lineages since the MRCA of living flowering plants. It was previously known that inner and outer integument development in *Arabidopsis* depended on



substantially distinct sets of molecular players. The current study provides evidence that most of these differences in molecular controls are ancient, and already existed in early flowering plants. The only clear exception noted in the present study, in which molecular mechanisms seem to have diverged in one or both of the *Amborella* and *Arabidopsis* lineages, concerns the paralogous ARF family members *ETT* and *ARF4*. *ETT* contributes to abaxial tissue identity uniquely in the inner integument of *Arabidopsis*, whereas orthologs of both *ETT* and *ARF4* are expressed in the abaxial domain of both integuments in *Amborella*. The inner integument expression of these factors may, as for that of *ETT* in *Arabidopsis*, act to prevent the coalescence of the two integument primordia. However, these factors may play a further role in the establishment and/or maintenance of abaxial/adaxial polarity in the *Amborella* outer integument, compared to their orthologs in *Arabidopsis*. Analysis of gene expression patterns in further ANA-grade taxa could be used to determine whether these ARF genes evolved more restricted inner integument-specific roles in the *Arabidopsis* lineage, or whether, conversely, their expression patterns became more generalized to both integuments, specifically in the *Amborella* lineage.

It is interesting to note that *ETT* expression is limited to the abaxial domain of the inner integument in bitegmic species of *Prunus* (rosid I clade) (Lora et al., 2015), as it is in *Arabidopsis*. If *ETT/ARF4* expression was lost from the outer integument in the *Arabidopsis* lineage, rather than gained in the *Amborella* lineage, it would therefore seem likely that this change happened before the divergence of the rosid I and II clades, which is estimated to have occurred some 105 MYA (Moore et al., 2010).

We have used the data presented here, together with information from published studies, to construct a partial molecular model for the control of integument development in the MRCA of living angiosperms (Figure 3). Predicted intermolecular interactions shown in this model are based on functional data from *Arabidopsis* alone and the conservation of tissue-colocalization of the molecules concerned in *Amborella*.

Implications of Reconstructed Integument Development Mechanisms for the Origin of the Outer Integument and Flowering Plants

Like the carpel, the outer integument is a pleisiomorphic feature of angiosperms. Two contrasting hypotheses have been proposed to account for the origin of this structure, prior to the radiation of living angiosperms. One of these hypotheses (reviewed by Doyle, 2008), based mainly on paleobotanical data, proposes that the angiosperms evolved from cupulate seed-ferns, possibly resembling the known fossil groups Caytoniales or Glossopteridales. In most known members of these taxa, several ovules occur within laminar cupules that are either borne on the margins of a female rachis (in Caytoniales), or emerge from the axil or midrib of a megasporophyll (in Glossopteridales). According to this “cupule hypothesis”, the bitegmic angiosperm ovule would have evolved from a unitegmic gymnosperm ovule by a reduction in the number of ovules-per-cupule to one, the

cupule thereby becoming the outer integument. The Caytoniales-type cupule may be a particularly strong candidate as a precursor to the angiosperm outer integument as the orientation of the ovule within this type of cupule could have led directly to the anatropous ovule arrangement (Doyle, 2008), which may have been pleisiomorphic in angiosperms.

The main alternative hypothesis for the origin of the outer integument, as discussed by Mathews and Kramer (2012), is that this structure simply arose through a modular reiteration of the single integument already present in species along the angiosperm stem lineage. Indeed, the development of the outer and inner integuments proceeds in a remarkably similar manner in most angiosperms, and the overexpression of *WUS* leads, in transgenic *Arabidopsis*, to the production of supernumerary integuments (Gross-Hardt et al., 2002; Sieber et al., 2004a), demonstrating that extra integuments can be generated from pre-existing developmental mechanisms by a simple molecular change in an upstream regulator. Not all strong candidates for ancestors, or close stem-lineage relatives of the angiosperms, possessed cupules. Notably, Bennettitales, which possessed many angiosperm-like characteristics, including a perianth, a bisexual axis (in some taxa), non-saccate pollen, net-veined leaves, and oleananes (highly resistant terpenoid compounds, also present in angiosperms), possessed no structure resembling an ovule-enclosing cupule. A mechanism based on the modular reiteration of a pre-existing single integument might, therefore, explain the origin of the angiosperm outer integument in a potential bennettitallian ancestor.

The data presented here strongly suggest that the precise mechanisms of integument development were already substantially distinct in the inner and outer integuments some 164 MYA in the MRCA of living flowering plants. This conclusion clearly favors the hypothesis of a distant or indirect homology between the inner and outer integuments, such as might be explained by the more direct origin of the outer integument from an ovule-containing seed-fern cupule. It does not provide support for the possible origin of the outer integument by a modular reiteration of the inner integument, as this mechanism would be expected to yield two integuments whose development initially depended, in early flowering plants, on near-identical sets of molecular regulators.

AUTHOR CONTRIBUTIONS

GA performed the experiments and phylogenetic analyses. AV supervised the experimental work and phylogenetic analyses. AA-R supervised the experimental work. BF and GG provided the plant material and contributed to writing the paper. CS planned and supervised the research and wrote the paper.

FUNDING

This research was supported by ANR grant “ORANGe” 2013-18 to CS. GA was supported by a normalien undergraduate scholarship, and AV was employed as an agrégée préparatrice (teacher/researcher) of the ENS-Lyon.

ACKNOWLEDGMENTS

We are grateful to the regional authorities of the province Nord and province Sud of New Caledonia for permission to collect living *Amborella* material.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.01352/full#supplementary-material>

FIGURE S1 | Maximum likelihood phylogeny of angiosperm HD-ZIP III family proteins. aLRT branch-support values are shown at nodes. *Amborella* and *Arabidopsis* sequences mentioned in the text are highlighted. Key to species names: *Ara tha*: *Arabidopsis thaliana*; *AmTr*: *Amborella trichopoda*; *Bra dis*: *Brachypodium distachyon*; *Bra nap*: *Brassica napus*; *Bra rapa*: *Brassica rapa*; *Cam sat*: *Camelina sativa*; *Cit sin*: *Citrus sinensis*; *Fra ves*: *Fragaria vesca*; *Gly max*: *Glycine max*; *Med tru*: *Medicago truncatula*; *Mus acu*: *Musa acuminata*; *Nel nuc*: *Nelumbo nucifera*; *Ory bra*: *Oryza brachyantha*; *Ory sat*: *Oryza sativa*; *Pop tri*: *Populus trichocarpa*; *Pru per*: *Prunus persica*; *Raph sat*: *Raphanus sativus*; *Ric com*: *Ricinus communis*; *Sol lyc*: *Solanum lycopersicum*; *Sol tub*: *Solanum tuberosum*; *The cac*: *Theobroma cacao*; *Vig ang*: *Vigna angularis*; *Vig rad*: *Vigna radiata*; *Vit vin*: *Vitis vinifera*; and *Zea may*: *Zea mays*.

REFERENCES

- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J. H., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389
- Bonaccorso, O., Lee, J. E., Puah, L., Scutt, C. P., and Golz, J. F. (2012). FILAMENTOUS FLOWER controls lateral organ development by acting as both an activator and a repressor. *BMC Plant Biol.* 12:176. doi: 10.1186/1471-2229-12-176
- Buzgo, M., Soltis, P. S., and Soltis, D. E. (2004). Floral developmental morphology of *Amborella trichopoda* (Amborellaceae). *Int. J. Plant Sci.* 165, 925–947. doi: 10.1086/424024
- Doyle, J. A. (2008). Integrating molecular phylogenetic and paleobotanical evidence on origin of the flower. *Int. J. Plant Sci.* 169, 816–843. doi: 10.1086/589887
- Endress, P. K. (2011). Angiosperm ovules: diversity, development, evolution. *Ann. Bot.* 107, 1465–1489. doi: 10.1093/aob/mcr120
- Endress, P. K., and Igersheim, A. (2000). Reproductive structures of the basal angiosperm *Amborella trichopoda* (Amborellaceae). *Int. J. Plant Sci.* 161, S237–S248. doi: 10.1086/317571
- Eshed, Y., Izhaki, A., Baum, S. F., Floyd, S. K., and Bowman, J. L. (2004). Asymmetric leaf development and blade expansion in *Arabidopsis* are mediated by KANADI and YABBY activities. *Development* 131, 2997–3006. doi: 10.1242/dev.01186
- Finet, C., Floyd, S. K., Conway, S. J., Zhong, B., Scutt, C. P., and Bowman, J. L. (2016). Evolution of the YABBY gene family in seed plants. *Evol. Dev.* 18, 116–126. doi: 10.1111/ede.12173
- Finet, C., Fourquin, C., Vinauger, M., Berne-Dedieu, A., Chambrier, P., Paindavoine, S., et al. (2010). Parallel structural evolution of auxin response factors in the angiosperms. *Plant J.* 63, 952–959. doi: 10.1111/j.1365-313X.2010.04292.x
- Fourquin, C., Vinauger-Douard, M., Fogliani, B., Dumas, C., and Scutt, C. P. (2005). Evidence that CRABS CLAW and TOUSLED have conserved their roles in carpel development since the ancestor of the extant angiosperms. *Proc. Natl. Acad. Sci. U.S.A.* 102, 4649–4654. doi: 10.1073/pnas.0409577102
- Gouy, M., Guindon, S., and Gascuel, O. (2010). SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* 27, 221–224. doi: 10.1093/molbev/msp259

tuberosum; *The cac*: *Theobroma cacao*; *Vig ang*: *Vigna angularis*; *Vig rad*: *Vigna radiata*; *Vit vin*: *Vitis vinifera*; and *Zea may*: *Zea mays*.

FIGURE S2 | Alignment of angiosperm HD-ZIP III family proteins used to produce the phylogeny shown in **Supplementary Figure 1**. Sites used in phylogenetic analysis are underlined.

FIGURE S3 | Maximum likelihood phylogeny of angiosperm KANADI family proteins. aLRT branch-support values are shown at nodes. *Amborella* and *Arabidopsis* sequences mentioned in the text are highlighted. Key to species names: *Ara tha*: *Arabidopsis thaliana*; *AmTr*: *Amborella trichopoda*; *Bra dis*: *Brachypodium distachyon*; *Bra nap*: *Brassica napus*; *Bra rapa*: *Brassica rapa*; *Cam sat*: *Camelina sativa*; *Cit sin*: *Citrus sinensis*; *Fra ves*: *Fragaria vesca*; *Gly max*: *Glycine max*; *Med tru*: *Medicago truncatula*; *Mus acu*: *Musa acuminata*; *Nel nuc*: *Nelumbo nucifera*; *Ory bra*: *Oryza brachyantha*; *Ory sat*: *Oryza sativa*; *Pop tri*: *Populus trichocarpa*; *Pru per*: *Prunus persica*; *Raph sat*: *Raphanus sativus*; *Ric com*: *Ricinus communis*; *Sol lyc*: *Solanum lycopersicum*; *Sol tub*: *Solanum tuberosum*; *The cac*: *Theobroma cacao*; *Vig ang*: *Vigna angularis*; *Vig rad*: *Vigna radiata*; *Vit vin*: *Vitis vinifera*; and *Zea may*: *Zea mays*.

FIGURE S4 | Alignment of angiosperm KANADI family proteins used to produce the phylogeny shown in **Supplementary Figure 3**. Sites used in phylogenetic analysis are underlined.

TABLE S1 | Oligonucleotide primers used to amplify *Amborella* integument regulators.

- Gross-Hardt, R., Lenhard, M., and Laux, T. (2002). WUSCHEL signaling functions in interregional communication during *Arabidopsis* ovule development. *Genes Dev.* 16, 1129–1138. doi: 10.1101/gad.225202
- Guindon, S., Dufayard, J.-F., Hordijk, W., Lefort, V., and Gascuel, O. (2009). PhyML: fast and accurate phylogeny reconstruction by maximum likelihood. *Infect. Genet. Evol.* 9, 384–385.
- Huang, T., Harrar, Y., Lin, C., Reinhart, B., Newell, N. R., Talavera-Rauh, F., et al. (2014). *Arabidopsis* KANADI1 acts as a transcriptional repressor by interacting with a specific cis-element and regulates auxin biosynthesis, transport, and signaling in opposition to HD-ZIP III factors. *Plant Cell* 26, 246–262. doi: 10.1105/tpc.113.111526
- Kelley, D. R., Arreola, A., Gallagher, T. L., and Gasser, C. S. (2012). ETTIN (ARF3) physically interacts with KANADI proteins to form a functional complex essential for integument development and polarity determination in *Arabidopsis*. *Development* 139, 1105–1109. doi: 10.1242/dev.067918
- Kelley, D. R., Skinner, D. J., and Gasser, C. S. (2009). Roles of polarity determinants in ovule development. *Plant J.* 57, 1054–1064. doi: 10.1111/j.1365-313X.2008.03752.x
- Linkies, A., Graeber, K., Knight, C., and Leubner-Metzger, G. (2010). The evolution of seeds. *New Phytol.* 186, 817–831. doi: 10.1111/j.1469-8137.2010.03249.x
- Lora, J., Hormaza, J. I., and Herrero, M. (2015). Transition from two to one integument in *Prunus* species: expression pattern of INNER NO OUTER (INO), ABERRANT TESTA SHAPE (ATS) and ETTIN (ETT). *New Phytol.* 208, 584–595. doi: 10.1111/nph.13460
- Mathews, S., and Kramer, E. M. (2012). The evolution of reproductive structures in seed plants: a re-examination based on insights from developmental genetics. *New Phytol.* 194, 910–923. doi: 10.1111/j.1469-8137.2012.04091.x
- McAbee, J. M., Hill, T. A., Skinner, D. J., Izhaki, A., Hauser, B. A., Meister, R. J., et al. (2006). ABERRANT TESTA SHAPE encodes a KANADI family member, linking polarity determination to separation and growth of *Arabidopsis* ovule integuments. *Plant J.* 46, 522–531. doi: 10.1111/j.1365-313X.2006.02717.x
- Moore, M. J., Soltis, P. S., Bell, C. D., Burleigh, J. G., and Soltis, D. E. (2010). Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4623–4628. doi: 10.1073/pnas.0907801107
- Nakata, M., Matsumoto, N., Tsugeki, R., Rikirsch, E., Laux, T., and Okada, K. (2012). Roles of the middle domain-specific WUSCHEL-RELATED HOMEBOX genes in early development of leaves in *Arabidopsis*. *Plant Cell* 24, 519–535. doi: 10.1105/tpc.111.092858

- Pekker, I., Alvarez, J. P., and Eshed, Y. (2005). Auxin response factors mediate *Arabidopsis* organ asymmetry via modulation of KANADI activity. *Plant Cell* 17, 2899–2910. doi: 10.1105/tpc.105.034876
- Reinhart, B. J., Liu, T., Newell, N. R., Magnani, E., Huang, T., Kerstetter, R., et al. (2013). Establishing a framework for the Ad/Abaxial regulatory network of *Arabidopsis*: ascertaining targets of class III HOMEODOMAIN LEUCINE ZIPPER and KANADI regulation. *Plant Cell* 25, 3228–3249. doi: 10.1105/tpc.113.111518
- Sieber, P., Gheyselinck, J., Gross-Hardt, R., Laux, T., Grossniklaus, U., and Schneitz, K. (2004a). Pattern formation during early ovule development in *Arabidopsis thaliana*. *Dev. Biol.* 273, 321–334. doi: 10.1016/j.ydbio.2004.05.037
- Sieber, P., Petrascheck, M., Barberis, A., and Schneitz, K. (2004b). Organ polarity in *Arabidopsis*. NOZZLE physically interacts with members of the YABBY family. *Plant Physiol.* 135, 2172–2185. doi: 10.1104/pp.104.040154
- Simon, M. K., Skinner, D. J., Gallagher, T. L., and Gasser, C. S. (2017). Integument development in *Arabidopsis* depends on interaction of YABBY protein INNER NO OUTER with coactivators and corepressors. *Genetics* 207, 1489–1500. doi: 10.1534/genetics.117.300140
- Stevens, P. F. (2001). *Angiosperm Phylogeny Website. Version 14*. Available at: <http://www.mobot.org/MOBOT/research/APweb/>
- Viallette-Guiraud, A. C., Alaux, M., Legeai, F., Finet, C., Chambrier, P., Brown, S. C., et al. (2011). Cabomba as a model for studies of early angiosperm evolution. *Ann. Bot.* 108, 589–598. doi: 10.1093/aob/mcr088
- Yamada, T., Ito, M., and Kato, M. (2003). Expression pattern of INNER NO OUTER homologue in *Nymphaea* (water lily family, Nymphaeaceae). *Dev. Genes Evol.* 213, 510–513. doi: 10.1007/s00427-003-0350-8
- Yamada, T., Sasaki, Y., Hashimoto, K., Nakajima, K., and Gasser, C. S. (2016). CORONA, PHABULOSA and PHAVOLUTA collaborate with BELL1 to confine WUSCHEL expression to the nucellus in *Arabidopsis* ovules. *Development* 143, 422–426. doi: 10.1242/dev.129833
- Yamada, T., Yokota, S., Hirayama, Y., Imaichi, R., Kato, M., and Gasser, C. S. (2011). Ancestral expression patterns and evolutionary diversification of YABBY genes in angiosperms. *Plant J.* 67, 26–36. doi: 10.1111/j.1365-313X.2011.04570.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 © 2018 Arnault, Viallette, Andres-Robin, Fogliani, Gâteblé and Scutt. 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.



Integrative Analysis of Three RNA Sequencing Methods Identifies Mutually Exclusive Exons of MADS-Box Isoforms During Early Bud Development in *Picea abies*

Shirin Akhter^{1†}, Warren W. Kretzschmar^{2†}, Veronika Nordal¹, Nicolas Delhomme³, Nathaniel R. Street⁴, Ove Nilsson³, Olof Emanuelsson² and Jens F. Sundström^{1*}

¹ Linnean Center for Plant Biology, Uppsala BioCenter, Department of Plant Biology, Swedish University of Agricultural Sciences, Uppsala, Sweden, ² Science for Life Laboratory, Department of Gene Technology, School of Engineering Sciences in Biotechnology, Chemistry and Health, KTH Royal Institute of Technology, Solna, Sweden, ³ Umeå Plant Science Centre, Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural Sciences, Umeå, Sweden, ⁴ Umeå Plant Science Centre, Department of Plant Physiology, Umeå University, Umeå, Sweden

OPEN ACCESS

Edited by:

Annette Becker,
Justus-Liebig-Universität Gießen,
Germany

Reviewed by:

Leonie Verhage,
CEA Grenoble, France
Günter Theißen,
Friedrich-Schiller-Universität Jena,
Germany

*Correspondence:

Jens F. Sundström
jens.sundstrom@slu.se

[†] Shared first authorship

Specialty section:

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

Received: 15 June 2018

Accepted: 18 October 2018

Published: 13 November 2018

Citation:

Akhter S, Kretzschmar WW, Nordal V, Delhomme N, Street NR, Nilsson O, Emanuelsson O and Sundström JF (2018) Integrative Analysis of Three RNA Sequencing Methods Identifies Mutually Exclusive Exons of MADS-Box Isoforms During Early Bud Development in *Picea abies*. *Front. Plant Sci.* 9:1625. doi: 10.3389/fpls.2018.01625

Recent efforts to sequence the genomes and transcriptomes of several gymnosperm species have revealed an increased complexity in certain gene families in gymnosperms as compared to angiosperms. One example of this is the gymnosperm sister clade to angiosperm TM3-like MADS-box genes, which at least in the conifer lineage has expanded in number of genes. We have previously identified a member of this sub-clade, the conifer gene *DEFICIENS AGAMOUS LIKE 19* (*DAL19*), as being specifically upregulated in cone-setting shoots. Here, we show through Sanger sequencing of mRNA-derived cDNA and mapping to assembled conifer genomic sequences that *DAL19* produces six mature mRNA splice variants in *Picea abies*. These splice variants use alternate first and last exons, while their four central exons constitute a core region present in all six transcripts. Thus, they are likely to be transcript isoforms. Quantitative Real-Time PCR revealed that two mutually exclusive first *DAL19* exons are differentially expressed across meristems that will form either male or female cones, or vegetative shoots. Furthermore, mRNA *in situ* hybridization revealed that two mutually exclusive last *DAL19* exons were expressed in a cell-specific pattern within bud meristems. Based on these findings in *DAL19*, we developed a sensitive approach to transcript isoform assembly from short-read sequencing of mRNA. We applied this method to 42 putative MADS-box core regions in *P. abies*, from which we assembled 1084 putative transcripts. We manually curated these transcripts to arrive at 933 assembled transcript isoforms of 38 putative MADS-box genes. 152 of these isoforms, which we assign to 28 putative MADS-box genes, were differentially expressed across eight female, male, and vegetative buds. We further provide evidence of the expression of 16 out of the 38 putative MADS-box genes by mapping PacBio Iso-Seq circular consensus reads derived from pooled sample sequencing to assembled transcripts. In summary, our

analyses reveal the use of mutually exclusive exons of MADS-box gene isoforms during early bud development in *P. abies*, and we find that the large number of identified MADS-box transcripts in *P. abies* results not only from expansion of the gene family through gene duplication events but also from the generation of numerous splice variants.

Keywords: *Picea abies*, MADS-box genes, cone development, De Bruijn assembly, transcript isoforms, RNA sequencing, *DAL19*

INTRODUCTION

In plants, members of the MADS-box gene family play important roles during diverse aspects of plant development and have been implicated in regulating e.g., floral transition and floral meristem and organ identity (see e.g., O'Maoileidigh et al., 2014 and references within). MADS is an acronym for the four founding members of this gene family: *MCM1* from *Saccharomyces cerevisiae*, *AGAMOUS* from *Arabidopsis thaliana*, *DEFICIENS* from *Antirrhinum majus*, and *SRF* from *Homo sapiens* (Schwarz-Sommer et al., 1990). MADS-box genes encode transcription factors. Evolutionary studies in seed plants have demonstrated that angiosperms and gymnosperms share orthologous MADS-box genes, and that these orthologs in many cases also are involved in similar biological processes. For a comprehensive review see Gramzow and Theissen (2010). For instance, MADS-box genes regulating carpel development in angiosperms have orthologous genes in gymnosperms involved in female cone development (Tandre et al., 1995; Winter et al., 1999).

We have previously produced inbred crosses of a naturally occurring spruce mutant *Picea abies* (var.) *acrocona* (Uddenberg et al., 2013). As reported, one quarter of the segregating sibling population resulting from those crosses initiated cones early, already during the second growth cycle, which suggests that genes of importance for vegetative to reproductive phase change are ectopically activated in the *acrocona* mutant. We have identified mature mRNA transcripts of the gene *DEFICIENS AGAMOUS LIKE 19* (*DAL19*) as being specifically upregulated in cone-setting shoots of early cone-setting *acrocona* plants (Uddenberg et al., 2013). *DAL19* belongs to the MIKC-type of MADS-box transcription factor proteins. MIKC refers to the different protein domains found in a class of plant-specific MADS-box genes (Gramzow and Theissen, 2010): M is short for the MADS-domain, which is responsible for DNA binding, I is a variable intervening region sometimes also referred to as the Linker, the K-domain is a keratin-like domain responsible for protein-protein interactions, and C is a variable C-terminal region (Ma et al., 1991), which in some MADS-box genes has been shown to encode activation domains (Litt and Irish, 2003).

Phylogenetic analyses have demonstrated that *DAL19*, together with closely related gymnosperm genes, form a distinct subclade in the MADS-box gene phylogeny (Uddenberg et al., 2013). The gymnosperm genes form a sister-clade to angiosperm *TM3*-like genes, which harbors the floral integrator *SUPPRESSOR OF OVEREXPRESSION OF CO 1* (*SOC1*), also known as *AGAMOUSLIKE20* (*AGL20*) from *Arabidopsis thaliana*

(Gramzow et al., 2014). *SOC1* and its orthologs in other angiosperm species, integrate several flowering signals derived from e.g., temperature, aging, plant hormones, and photoperiod to regulate the transition from vegetative to reproductive phase (Dorca-Fornell et al., 2011), and (Gramzow et al., 2014) hypothesize that at least some members in the gymnosperm *TM3* sister-clade perform similar functions.

The *DAL19* transcripts were first cloned by 3' and 5' Rapid Amplification of cDNA Ends (RACE), and by mRNA *in situ* hybridization shown to be preferentially expressed in the abaxial part of the ovuliferous scales in female cones (Carlsbecker et al., 2013). Sequence comparison between the *DAL19* clone (KC347015) presented in Carlsbecker et al. (2013) and the assembled *DAL19* transcript presented in Uddenberg et al. (2013) (*Acr42124_1*) revealed distinct sequence differences in the 5' part of the transcript that are all contained within the first exon.

Mortazavi et al. (2008) showed that massively parallel short-read mRNA sequencing could be used together with a well-annotated reference genome to map reads across splice junctions to identify novel transcript isoforms in mouse and human. Later studies in plants employed reference-guided assembly of short-read mRNA sequencing to annotate reference genomes with new alternatively spliced transcript isoforms in *Arabidopsis thaliana* (Marquez et al., 2012), *Zea mays* (Thatcher et al., 2014), and *Brachypodium distachyon* (Mandadi and Scholthof, 2015). However, in organisms where high-quality reference genomes are not available, *de novo* transcriptome assembly (Robertson et al., 2010; Grabherr et al., 2011; Schulz et al., 2012) is the only approach available for transcript isoform reconstruction from short reads (Conesa et al., 2016). Unfortunately, current methods for *de novo* transcriptome assembly have difficulties reconstructing full-length transcript isoforms (Steijger et al., 2013; Xu et al., 2015; Conesa et al., 2016). Several recent plant studies adopted a hybrid sequencing approach where long, error-corrected reads derived from Pacific Biosciences' isoform sequencing technology, Iso-Seq, were used to identify novel transcripts, and short, high-quality reads from massively parallel mRNA sequencing were used to correct and quantify those transcripts by reference-guided assembly (Xu et al., 2015; Abdel-Ghany et al., 2016; Wang et al., 2016; Liu et al., 2017) or by reference-free assembly (Hoang et al., 2017).

Hybrid sequencing studies of organisms with a well-annotated reference genome found evidence that the inclusion of Iso-Seq reads increased the sensitivity to alternatively splicing transcript isoforms. In root tissue of *Salvia miltiorrhiza*, Xu et al. (2015) reported detecting 110,715 and 1,109,011 known splice junctions with Illumina short reads and Iso-Seq, respectively. Using just Iso-Seq reads, Abdel-Ghany et al. (2016) increased the number

of known transcript isoforms in sorghum from 2,950 to 10,053, and Wang et al. (2016) increased the number of known transcript isoforms in *Zea mays* from 63,540 to 111,151. However, when compared on a transcriptome coverage basis, long reads from Iso-Seq are substantially more expensive than Illumina short reads, and this is reflected in the lack of long-read sequencing depth to identify uncommon transcripts in both Liu et al. (2017) and Hoang et al. (2017). For example, Liu et al. (2017) found that their Iso-Seq data only covered 62% of the multi-exonic genes in their chosen reference gene model, and Hoang et al. (2017) found that only 87% of Illumina short reads mapped back to transcript isoforms derived from Iso-Seq.

In this study, we only had Iso-Seq reads from pooled sample sequencing available, but our interest was in exploring the transcript isoforms in a subset of those samples for which we had short reads from total mRNA sequencing (RNA-seq). Lacking a finished reference genome for *P. abies* we tried assembling our short reads with Trinity (Grabherr et al., 2011) and Oases (Schulz et al., 2012), two popular assemblers built around De Bruijn graphs. Unfortunately, these methods did not reconstruct *DAL19* isoforms well, so we developed a novel approach that combines naive assembly from a De Bruijn graph with kallisto (Bray et al., 2016). We use this approach to construct a parsimonious set of full-length transcript isoforms for *DAL19*.

We verify the presence of multiple mature mRNA transcript isoforms of *DAL19* using three independent methods: (i) Rapid Amplification of cDNA Ends (RACE) followed by Sanger sequencing (Yeku and Frohman, 2011), (ii) a novel method for *de novo* assembly of short RNA-seq reads *in silico*, and (iii) PacBio Iso-Seq RNA sequencing (Sharon et al., 2013). Furthermore, we map the *DAL19* transcripts derived from Sanger sequencing to the genome sequences of *P. abies* and the closely related *Picea glauca* to address whether these *DAL19* transcripts are transcribed from one single locus, and hence can be considered transcript isoforms, or if these transcripts are transcribed from multiple loci.

In *P. abies*, female and male cones and vegetative shoots are initiated as buds on the growing shoots, i.e., female and male cones are formed from separate meristems (Hannerz and Sundström, 2006). Differential expression of the different isoforms in these tissues would be evidence that these isoforms are biologically relevant. Therefore, to address if the *DAL19* variants play distinct roles during development, we assay their transcription in young developing buds using quantitative Real-Time PCR (qRT-PCR) and mRNA *in situ* hybridization methods.

Finally, we address whether the observed pattern of multiple splice variants of *DAL19* is also present in other members of the MADS-box gene family in *P. abies*. Using our novel method for *de novo* assembly of short RNA-seq reads *in silico* and using long-read PacBio Iso-Seq RNA-seq data, we provide evidence for extensive alternative splicing of mutually exclusive exons among *P. abies* MADS-box genes that form a sister clade to angiosperm *TM3*-like genes. These findings suggest that vegetative to reproductive phase change and cone setting in *P. abies* are regulated by a genetic mechanism fine-tuned by the usage of different MADS-box isoforms in different bud types.

MATERIALS AND METHODS

Plant Material

Plant material was collected from an adult tree of Norway spruce (*Picea abies* L. Karst.) at the Rörby seed orchard (latitude 59°54' N) outside of Uppsala, Sweden. Male, female, and vegetative buds were collected during two distinct developmental phases of early bud development: (i) meristematic buds before or at the onset of lateral organ formation, (ii) buds that had progressed into active lateral organ formation and cell differentiation. Meristematic buds were collected on August 1st 2016. Lateral organ initiating buds were collected on August 16th 2015 or on September 12th 2016. In our study, a biological replicate consisted of a pool of three buds of one bud type from one adult tree. All plant materials used for RNA preparations were snap-frozen in liquid nitrogen and stored at -70°C. Plant materials used for *in situ* hybridization experiments were directly collected into fixative media according to Karlgren et al. (2009).

Samples for PacBio Iso-Seq sequencing were collected from several tissue types and conditions to give a broad representation of transcripts expressed in *Picea abies*. In total, 33 samples were collected and the tissue types included: (i) developmental samples of roots, hypocotyl, SAM and cotyledons, vegetative-, male- and female buds, pollen, ovuliferous scales and ovules, (ii) diurnal samples of cambial tissue, (iii) vascular samples of xylem and phloem, and (iv) cold stress samples of roots. All samples collected for PacBio Iso-Seq sequencing were snap-frozen in liquid nitrogen and stored at -70°C. Total RNA was prepared as described below (see RNA Preparation) and the resulting RNA-samples were pooled before performing PacBio Iso-Seq sequencing (see RNA Sequencing). We describe the samples in **Supplementary Table S1**.

RNA Preparation

Tissue homogenization, extraction, CHISAM (chloroform/isoamylalcohol 24:1) purification and isopropanol precipitation CHhomoRNA were performed as described by Azevedo et al. (2003). Resulting RNA pellets were dissolved in 350 µl RLT buffer (Qiagen RNeasy Kit) and long mRNAs were separated from microRNAs according to manufacturer's instructions (Qiagen RNeasy Kit, Qiagen, Carlsbad, CA, United States).

Cloning of *DAL19* Transcripts

Alternative mature mRNA transcript ends of *DAL19* were verified by synthesizing cDNA ends from mixed bud tissues (male, female and vegetative) and performing 5' RACE and 3' RACE according to the manufacturer protocol (FirstChoice® RLM-RACE Kit, Thermo Fisher Scientific, Vilnius, Lithuania). Primers that were used in this approach are listed in **Supplementary Tables S2B,C**.

Isolated alternative cDNA ends from 5' RACE and 3' RACE were cloned into the PCR blunt II Topo vector and transformed into *Escherichia coli* according to manufacturer instructions (Zero Blunt® TOPO® PCR Cloning Kit, Invitrogen, Carlsbad, CA, United States). Transformants that carried the correct insert were selected using colony-PCR with gene specific primers and

RACE primers provided by the kit listed in **Supplementary Table S2**. Plasmids were prepared using the GeneZet plasmid mini prep kit according to manufacturer instruction (Thermo Fisher Scientific, Vilnius, Lithuania). All plasmids were sent to GATC Biotech, Germany, for sequence verification.

Quantitative Real-Time PCR

qRT-PCR was performed using CFX-connect Real-Time System on 96-well PCR plates with adhesive seals (Bio-Rad Laboratories). Expression was analyzed in three distinct biological samples of each tissue type, except for vegetative buds collected on September 12th 2016 in which one sample was lost due to RNA degradation. Each biological sample was analyzed in triplicate. Primers used to quantify expression levels are provided in **Supplementary Table S2D**. The expression data of each gene was normalized against the expression of three reference genes, *ACTIN*, *POLYUBIQUITIN*, and *HISTONE2A*. PCR amplifications were carried out using the Maxima SYBR Green/ROX qPCR kit (ThermoScientific, Vilnius, Lithuania). Total 10 ng cDNA was used in 25 μ l PCR reaction. PCR cycling conditions were followed by the manufacturer instructions. The reactions were run for 40 cycles. Melt curves were generated to ensure product uniformity at the end of each run. Inter-run connector samples were included in all studies to correct for the use of multiple plates. Calculations and normalizations were done using the CFX software based on the “ Δ Ct or $\Delta\Delta$ Ct methods” (Bio-Rad).

Statistical analyses were performed using R 3.4.2. Normalized gene expression values from qRT-PCR experiments were used as an input dataset in the analyses. Student's *t*-test was performed to assess the significance of differences of gene expression between samples.

In situ Hybridization

Tissue fixation, embedding, sectioning, *in situ* section pre-treatment, and preparation of hybridization solutions were performed as described by Karlgren et al. (2009). LNA probes with complementary sequences to *DAL19* variants were synthesized, 5' and 3' digoxigenin-labeled (Exiqon) (**Supplementary Table S2E**), and then hybridized at 52°C. *In situ* post hybridization was performed as described previously (Karlgren et al., 2009). The images of hybridization signals were taken using Zeiss Axioplan microscope equipped with an AxioCam ICc 5 camera.

RNA Sequencing

2 \times 125 bp short-read paired-end RNA sequencing of mRNA-enriched male, female, and vegetative bud samples collected on August 1st 2016 was performed using a HiSeq2500 with v4-sequencing chemistry by The SNP & SEQ Technology Platform in Uppsala, Sweden. We pre-processed all reads by removing ribosomal RNA with SortMeRNA version 2.1b (Kopylova et al., 2012), and by removing adapters and low-quality parts of reads with Trimmomatic version 0.36 (Bolger et al., 2014). The complete command can be found in **Supplementary File S1**.

The SNP & SEQ Technology Platform in Uppsala, Sweden, performed PacBio Iso-Seq sequencing of the pooled samples on

a PacBio Sequel System. They processed the raw reads with the SMRTLink v5.0.1 Iso-Seq workflow to obtain circular consensus (CCS) reads. We present the subset of reads associated with *DAL1-DAL41*. Remaining reads will be made available at the conifer genomic resource database¹.

Discovery of Putative MADS-Box Core Region Sequences

We compiled a list of 32 previously published *P. abies* MADS-box genes (Carlsbecker et al., 2013) as well as ten newly identified full-length MIKC MADS-box genes retrieved from PlantGenIE.org (Sundell et al., 2015), here denoted *DAL31-DAL41* (**Supplementary Table S3**). We identified the full length MIKC MADS-box genes by performing BLAST searches using the option BLASTN- nucleotide query to nucleotide db and selecting blast DB *P. abies* 1.0 (complete) (Sundell et al., 2015). BLAST searches resulted in several scaffolds aligning to each cDNA sequence. We identified conserved intron/exon borders in the cDNA sequences by comparing the exon organization of spruce MADS-box genes to each other. The position of the first conserved intron/exon border is situated in the 3' region of the MADS-domain. Similarly, we found a conserved intron/exon border that exists at the end of the K-domain of each MADS-box gene of *P. abies*. We defined core region sequences of *P. abies* MADS-box genes as starting at the first conserved intron/exon border and ending at the last conserved intron/exon border of the K-domain (**Supplementary File S2**).

Assembly, Evaluation, and Visualization of *DAL19* Transcripts

We tested and developed RNA-seq assembly methods on a single training sample of vegetative *P. abies* buds. We created the Trinity (Grabherr et al., 2011) assembly from all reads that had a mate after trimming with Trimmomatic using Trinity version 2.4.0 (see **Supplementary Listing S1**). We created an Oases (Schulz et al., 2012) assembly from trimmed reads that had been merged with Pear version 0.9.10 (Zhang et al., 2014) using the “--keep-original” flag, and using Velvet 1.2.10 with Oases version 0.2.09.

We created an assembly using our own method: With Mccortex version v0.0.3-554-ga7d6f3b (Turner et al., 2018), we built and inferred a colored De Bruijn graph of *k*-mer size 47 from all reads that had a mate after trimming with Trimmomatic. We collapsed maximal runs of adjacent nodes with one incoming and one outgoing edge, excepting nodes at the ends of the runs, into unitigs. We removed unitigs with a mean coverage below four from the graph. We then traversed the resulting Cortex graph from every *k*-mer in *DAL19_ψ*, which resulted in one subgraph. We removed tips shorter than 47 *k*-mers from this subgraph. We created candidate transcripts from this cleaned graph by traversing every simple path from every incoming tip to every outgoing tip. Simple paths cannot contain cycles and therefore our candidate transcripts do not contain cycles either. For every sample, we assessed read support for all candidate transcripts with kallisto version 0.43.1 (Bray et al., 2016) from all reads that

¹congenie.org

had a mate after trimming. We ran kallisto with 100 bootstrap samples. We kept for further analysis all candidate transcripts for which the number of estimated counts was one or more in at least 95 bootstrap samples.

We evaluated each transcriptome assembly method based on the length of the match of the transcript with the maximal match to each *DAL19* transcript isoform. We calculated the maximal match length for every assembly by mapping all reads to the four *DAL19* transcript isoforms $\alpha\psi\gamma$, $\alpha\psi\delta$, $\beta\psi\gamma$, and $\beta\psi\delta$ with Minimap2 (default parameters) version 2.4-r555 (Li, 2018). See **Figure 1C** for a description of these isoforms. Match length was calculated for each transcript by counting the number of alignment matches in the alignment CIGAR string (SAMv1 specification²) and subtracting the number of base mismatches in the NM tag. A similarity proportion was calculated by dividing the length of the target *DAL19* transcript by the assembled transcript match length to that *DAL19* transcript.

We mapped all CCS reads obtained from the pooled sample to 42 core sequences of putative MADS-box genes in *P. abies* using Minimap2 version 2.4-r555 (Li, 2018) with the non-default argument “-x map-pb”. Twelve CCS reads had a primary alignment to one of the *DAL19* regions α , β , γ , δ , and ψ . We converted these reads to separate Cortex graphs of k -mer size 47. For each CCS-read graph, we joined the graph with the short-read graph and pruned unitigs with coverage below two. We then aggregated the twelve CCS reads by whether they had a primary or supplementary alignment to *DAL19* $_{\alpha}$, *DAL19* $_{\beta}$, or neither of these two regions. Four CCS reads fell in each of those categories. We merged the four CCS-read graphs of each *DAL19*-region-based group into a single color. We joined the resulting three graphs with the short-read graph and five graphs consisting of the k -mers of the *DAL19* regions α , β , γ , δ , and ψ .

For the purpose of visualization, the untig-filtered and tip-pruned Cortex graph was explored by traversing every connecting k -mer of any color from the initial 47-mer of *DAL19* $_{\psi}$ using cortexpy version 0.23.5.³ The complete command can be found in **Supplementary File S1**. We visualized the graph traversal of the final Cortex graph (**Figure 2**) with the Javascript library D3 version 4 (Jasaitiene et al., 2011) and Circos (Krzywinski et al., 2009). We calculated the graph layout with a Javascript implementation of CoLa (Dwyer et al., 2006).

Assembly and Filtering of MADS-Box Transcripts

For each of nine meristematic bud samples collected from a spruce tree on August 1st 2016 (three pools of male, vegetative, and female buds), we assembled transcript isoforms from pre-processed reads using our method (see Assembly, Evaluation, and Visualization of *DAL19* Transcripts) with the difference that we traversed the unitig-filtered Cortex graph from every k -mer in the MADS-core sequences (see Discovery of Putative MADS-Box Core Region Sequences), which resulted in several subgraphs. For each subgraph, we removed tips shorter than 47 k -mers, created candidate transcripts, and filtered those transcripts with kallisto

as described in Section “Assembly, Evaluation, and Visualization of *DAL19* Transcripts.” We combined all transcripts assembled in this way across all nine samples in a single file while keeping a single copy of duplicate transcripts. Transcripts that are reverse complements of each other were considered to be duplicates for the purpose of this candidate transcript aggregation step. The resulting file contained 1084 sequences.

We annotated the combined set of assembled transcripts with putative protein domains using Translated Reverse Position Specific BLAST (“rpstblastn” from the “blast” package version 2.6.0, build Jan 2 2018) against the NCBI’s Conserved Domain Database (revised on 16 January 2015) (Marchler-Bauer et al., 2015) using an e -value threshold of 0.01. We kept for further analysis 1073 MADS-box transcript sequences that could be annotated with at least one protein domain.

Assignment of Assembled MADS-Box Transcripts to MADS-Box Core Regions

We clustered the set of 1073 assembled transcripts that could be annotated with at least one protein domain using CD-HIT-EST version 4.6.4 (Li and Godzik, 2006) with the tuning parameters “-c 0.95 -n 10” in order to obtain clusters with sequence similarity above 95%. CD-HIT-EST provides a representative sequence for each cluster, and we mapped every cluster representative to our list of MADS-box core sequences using BWA-MEM version 0.7.8-r455 with default parameters. Finally, we assigned each assembled transcript in a cluster to the MADS-box core sequence that was the primary mapping of the cluster’s representative sequence. Thousand and forty assembled transcripts from 124 clusters could be assigned to 29 MADS-box core sequences in this way.

Transcript Isoform Curation and Multiple Sequence Alignment

The 1040 assembled transcripts assigned to a specific MADS-box core region were translationally aligned using the ClustalW module within Geneious (Geneious pro version 10.2.3 created by Biomatters Ltd.)⁴. Using the alignment, transcript isoforms were identified by manual curation. BLASTN searches (Scoring matrix: BLOSUM62; E -value threshold: $1e-3$) against the *P. abies* genome assembly V1.0⁵ were used to assign scaffolds from the genome assembly that contain 5’ or 3’ exons of MADS-box genes to each transcript isoform.

Mapping of PacBio Circular Consensus Reads to MADS-Box Core Regions

Minimap2 version 2.4-r555 (Li, 2018) using the preset “map-pb” was used to map circular consensus (CCS) PacBio Iso-Seq reads to the assembled and filtered transcripts. Only alignments with an edit distance less than five were kept. The assignment of transcripts to MADS-box scaffolds described in Section “Assembly, Evaluation, and Visualization of *DAL19* Transcripts.” was used to assign every CCS alignment to a MADS-box scaffold.

²<https://github.com/samtools/hts-specs>

³<https://github.com/winni2k/cortexpy>

⁴<http://www.geneious.com>

⁵<http://congenie.org>

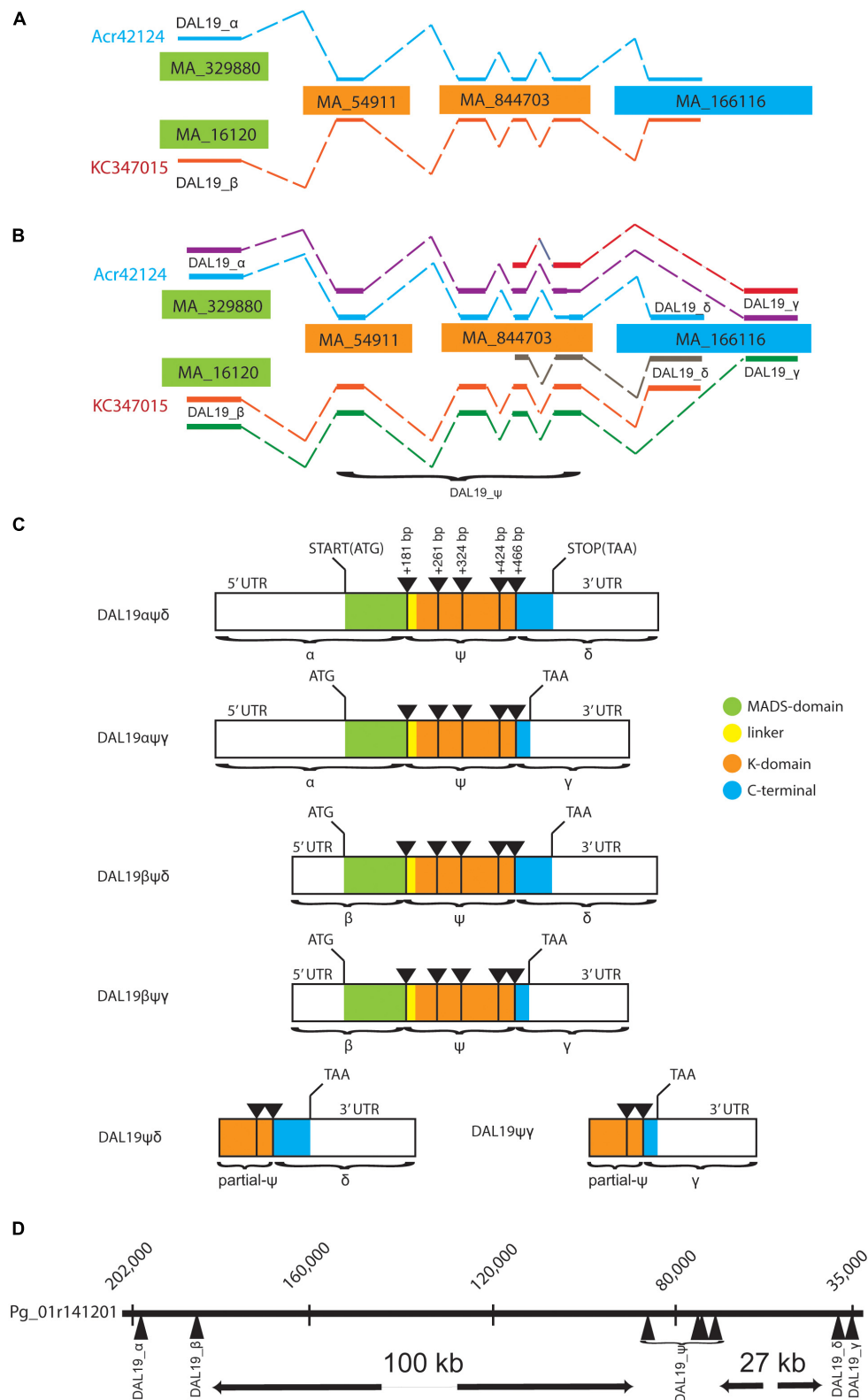


FIGURE 1 | Schematic view of *DAL19* isoforms and their genomic organization. **(A)** The two transcripts Acr42124 and KC347015 mapped to *Picea abies* (V1) genome scaffolds. **(B)** The additional transcripts presented in the current manuscript are included. The *DAL19* exons α, β, δ, and γ, and the core region ψ are indicated. **(C)** Illustration of the cloned *DAL19* isoforms. The MIKC-protein domain structure is color indicated, as well as the position of 5' and 3' UTRs, translational start and stop codons, and the position of intron/exon borders. **(D)** *DAL19* exons mapped to the *Picea glauca* genomic scaffold Pg01r141201.

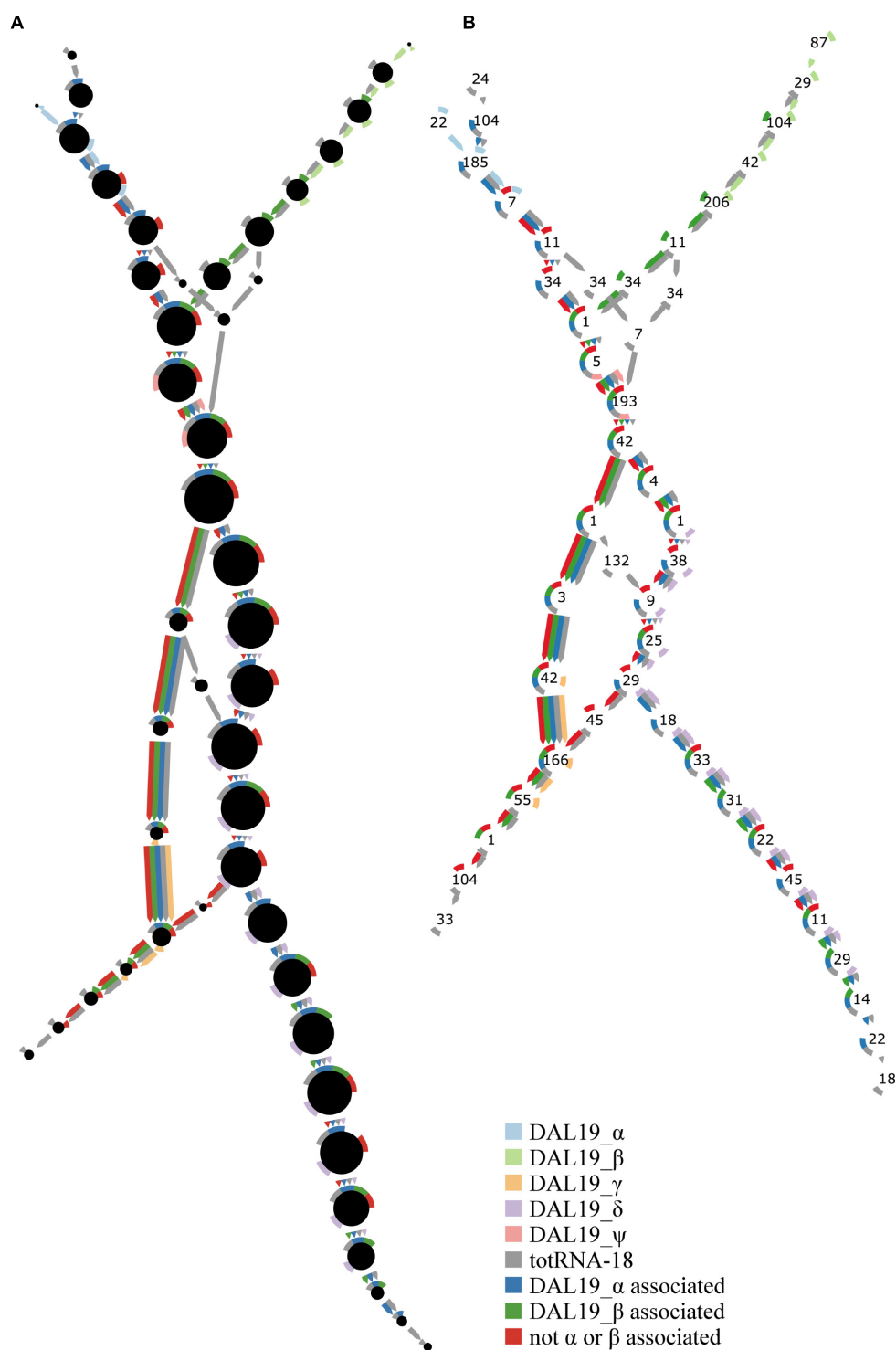


FIGURE 2 | Complete traversal of final Cortex graph using all colors and starting from the first 47-mer of *DAL19_ψ*. **(A,B)** Represent the same graph topology. Colors represent different sources of sequencing data as defined in the figure. Circles represent units. The rim of each circle consists of nine circular arcs of equal size representing one of each k-mer color. An arc is colored if the circle's unitig has non-zero coverage of k-mers in that color. A colored edge connecting circles represents an edge between end k-mers of a unitig in the color of the edge. The first five light colors represent the *DAL19* exons α, β, γ, and δ, and the core region ψ (obtained from cloning followed by Sanger sequencing). Gray represents k-mers from short-read RNA-seq data of "totRNA18." Dark blue represents four CCS reads that map to *DAL19_α*. Dark green represents four CCS reads that map to *DAL19_β*. Dark red represents four CCS reads that map to *DAL19_ψ*, but not the α or β region. **(A)** The area of the black central circle is proportional to the maximum of the per-color mean coverage of the unitig represented by that circle. **(B)** The number in the center of each circle is the number of k-mers that constitute the unitig represented by that circle.

Reads for which all primary and secondary alignments of a CCS read aggregated to the same MADS-box scaffold were counted in **Table 1**.

TABLE 1 | Long read and differential expression analysis results aggregated by scaffold (MA ID).

Gene name	MA_ID	Number of uniquely mapping CCS reads	Number of significant DE transcripts	Number of transcripts
DAL1	MA_128164	20	15	80
DAL1	MA_369170	0	0	6
DAL10	MA_15122short	0	0	6
DAL10	MA_15122long	0	25	78
DAL11	MA_18048short	0	0	7
DAL11	MA_18048long	0	6	15
DAL12	MA_13933short	0	4	8
DAL12	MA_13933long	0	12	20
DAL13	MA_130755	7	9	25
DAL13_lik_a	MA_2191711	0	2	4
DAL13_lik_b	MA_8198323	0	0	3
DAL19	MA_16120	3	1	20
DAL19	MA_162822_long	0	13	47
DAL19	MA_329880	2	2	22
DAL19	MA_162822short	0	1	1
DAL21	MA_126898long	0	4	17
DAL21	MA_126898short	0	0	4
DAL3	MA_141872	10	6	23
DAL3	MA_10256834	0	6	32
DAL3_lik_a	MA_19007	0	1	17
DAL3_lik_a	MA_32676	1	1	16
DAL31	MA_46639	7	2	22
DAL32	MA_7858089	0	2	16
DAL32	MA_138440long	2	5	39
DAL32	MA_138440short	0	0	4
DAL33	MA_10079394	0	13	92
DAL33	MA_276701	3	1	31
DAL35	MA_9284799	1	3	20
DAL37	MA_18942	10	0	20
DAL38	A_20180long	6	1	11
DAL38	A_20180short	0	0	1
DAL39	MA_94049	0	4	25
DAL4	MA_101463	9	2	110
DAL4	MA_35712	0	1	29
DAL40/JTL	MA_10432602	3	0	14
DAL41	MA_57186	1	7	26
DAL5	MA_32019	0	0	3
DAL9	MA_6279308	5	3	19
Sum		90	152	933

MA_ID: scaffold identification in *P. abies* genome assembly v1.0. Number of uniquely mapping CCS reads: Counts of long reads mapping uniquely to transcripts associated with a single scaffold. Number of significant DE transcripts: Number of significant associations at FDR 0.1 of transcript differential expression in eight bud samples, across bud types, and aggregated by MA ID. Of 38 groups of transcripts assigned to a single MA ID, 28 groups contained at least one transcript with differential expression across vegetative, male, and female buds at FDR 0.1. Number of transcripts: Number of transcripts assigned to each MA ID.

Estimation of Transcript Abundance With Kallisto

For each sample, we mapped read pairs processed only with SortMeRNA and Trimmomatic to the combined set of assembled transcripts using BWA-MEM version 0.7.8-r455 (Li, 2013) with default parameters and a minimum seed length of 47. We estimated the abundance of assembled transcripts with kallisto version 0.43.1 (default parameters) using only read pairs for which both reads mapped to any assembled transcript.

Differential Expression Analysis

We converted estimated transcript counts calculated by kallisto to read counts using the tximport package (version 1.8.0) (Soneson et al., 2015). We compared samples based on all transcripts with a maximum read count greater than one (823 transcripts) by (i) calculating Spearman correlation coefficients of the raw transcript counts (**Supplementary Figure S1**), by (ii) creating a heatmap of log-scaled transcript read counts (**Supplementary Figure S2**), and by (iii) performing multi-dimensional scaling (Euclidean distance) (**Supplementary Figure S3**). Heatmaps were created using the function heatmap.2 from the gplots package (version 3.0.1, [warnes2016]) in R (version 3.5.0, [r_core_team]). In **Supplementary Figures S1–S3** one male sample looks substantially different from the other two male samples. We removed this sample (“A43_S6_L001”) from the differential expression analysis of the assembled transcripts of nine bud samples.

We performed differential expression analysis on the converted read counts using a negative binomial Generalized Linear Model implemented in edgeR (version 3.22.1, R version 3.5.0) (Robinson et al., 2010) as a quasi-likelihood F-test (McCarthy et al., 2012) for any differences between all three vegetative bud groups. We used the Benjamini–Hochberg (Benjamini and Hochberg, 1995) procedure to correct *p*-values for false discovery rate. The results of differential expression analysis can be found in **Supplementary Table S6**.

Phylogenetic Analyses

Two separate analyses were performed. In the first, the nucleotide sequence corresponding to the complete open reading frame (ORF) of representative transcript isoforms were translationally aligned using the ClustalW module within Geneious to annotated MADS-box genes retrieved from GenBank from the gymnosperms *Picea abies* and *Pinus radiata* and the angiosperms *Arabidopsis thaliana* and *Lycopersicon esculentum* (**Supplementary Table S4** and **Supplementary Files S3, S4**). Phylogenetic parsimony analysis of the sequences (in total 116 taxa and 1385 characters) was performed using PAUP 4.0 a (build 161) (Swofford, 2002). A heuristic search with 1000 RANDOM additions was performed with the tree-bisection reconnection (TBR), steepest descent, and MULPARS options in effect. Support for the different groups in the tree was estimated using 100 bootstrap samples using the same heuristic search settings. In the second analysis, sequences outside of the MADS-box were trimmed away. The alignment of the MADS-box and subsequent parsimony analysis (120 taxa and 190

characters) was performed using the same settings as for the full length ORF.

RESULTS

Cloning of *DAL19* Transcripts Reveals Four Novel Transcript Isoforms

In order to verify the presence of the different *DAL19* mature mRNA transcripts we first mapped the previously published *DAL19* transcripts, Acr42124 and KC347015, to the published *Picea abies* genome sequence (*Picea abies* v 1.0) (Nystedt et al., 2013). Due to the large genome size and prevalence of highly repetitive regions, scaffolds in the current genome assembly of *Picea abies* are relatively short and genes are commonly scattered over several genomic scaffolds. In line with this, both Acr42124 and KC347015 mapped to multiple genomic scaffolds (Figure 1A). Notably the sequence corresponding to the first exon of Acr42124 mapped to the genomic scaffold MA_329880 of 9.4 kbp, whereas the first exon of KC347015 mapped to the MA_16120 scaffold of 74 kbp (Figure 1B). We named these alternate exons *DAL19* α and *DAL19* β , respectively. The remaining exons mapped to the same set of genomic scaffolds: MA_54911, MA_844703, and MA_166116. In order to connect *DAL19* α and *DAL19* β with the adjacent exon on scaffold MA_54911, forward primers placed in the 5' untranslated regions of either *DAL19* α or *DAL19* β were used in combination with reversed primers placed in the adjacent exon to amplify partial *DAL19* transcripts starting with either the *DAL19* α or *DAL19* β exons. Next, we used gene-specific *DAL19* primers in combination with generic 3' poly-T or 5'CAP primers to amplify the 3' or 5' ends of *DAL19*. Surprisingly, this resulted in the amplification of a version of *DAL19* with an alternate 3' end as compared to the previously published 3' ends of KC347015 and Acr42124. The sequence that differs corresponds to the last exon of *DAL19* (Figure 1B). Both exons (henceforth called *DAL19* δ and *DAL19* γ , respectively) map to the same genomic scaffold (MA_166116) but at different positions. *DAL19* δ maps to position 4650–4944 and harbors the C-terminal signature motif of the TM3-clade EVETQL, whereas *DAL19* γ maps to position 3239–3505 and harbors a premature stop codon. In addition, the 5' RACE yielded two short versions of *DAL19* that lacked the MADS-box altogether. In summary, six versions of *DAL19* have been cloned: four long versions with alternate first and last exons (*DAL19* $\alpha\psi\delta$, *DAL19* $\alpha\psi\gamma$, *DAL19* $\beta\psi\delta$, and *DAL19* $\beta\psi\gamma$), and two short versions (*DAL19* $\psi\delta$ and *DAL19* $\psi\gamma$), where ψ refers to a common core region with four exons, as detailed in Figure 1C. All versions have been independently cloned in their full-length versions and sequenced by Sanger sequencing.

Mapping of *DAL19* Transcripts to Genomic Assemblies Provides Support That These Transcripts Are Isoforms

Given the fragmented assembly of the *P. abies* genome, it is not possible to rule out that we have identified *DAL19*

transcripts from different genomic loci. However, the presence of both *DAL19* δ and *DAL19* γ on the same genomic scaffold, and the fact that *DAL19* α and *DAL19* β are expressed together with either *DAL19* δ or *DAL19* γ , supports the notion that the *DAL19* transcripts we observed are indeed transcribed from one large locus. To provide independent support for this hypothesis we mapped the cloned *DAL19* transcripts to the genomic assembly of *Picea glauca* (PG29-V4.0), with which *P. abies* shares sequence similarity both in terms of frequent synteny and sequence identity (Sundell et al., 2015). The *DAL19* transcripts all mapped to the same *Picea glauca* scaffold (Figure 1D). According to the *P. glauca* assembly, the two variants of the first exon of Pg*DAL19* (Pg*DAL19* α and Pg*DAL19* β) are located approximately 16 kb apart and separated from the next *DAL19* exon by an intron of approximately 100 kb. The *DAL19* α exon is located 5' of the *DAL19* β exon, which possibly has its own promoter inside the large first intron of *DAL19*. The introns between exons three, four, and five of *DAL19* are relatively short and consist of only approximately one hundred bases each. We consider this region the core region ψ of the *DAL19* gene (or *DAL19* ψ) because the fifth exon is followed by a long intron of approximately 28 kb separating this region of *DAL19* from the two alternative 3' exons *DAL19* δ and *DAL19* γ .

Transcripts From Long-Read Sequencing and a Novel Transcriptome Assembly Method Are Consistent With Four *DAL19* Transcript Isoforms

In order to provide independent evidence of all cloned *DAL19* transcripts, we obtained short-read Illumina and long-read PacBio Iso-Seq circular consensus sequences (CCS) of total, long (>200 base pairs), polyadenylated RNA derived from a single sample of *P. abies* vegetative buds, and a pool of 33 *P. abies* samples, respectively. We represented the short reads as a De Bruijn graph, and then threaded the long reads through this graph. We then fully traversed the graph starting from the *DAL19* core region (*DAL19* ψ). This traversal (Figure 2) shows short-read RNA-seq *k*-mer coverage for the α , β , γ , δ , and ψ regions of *DAL19*. Short-read coverage is split evenly between the α and β regions. When the α and β regions meet at *DAL19* ψ , the short-read coverage merges additively. At the end of *DAL19* ψ , the majority of short-read coverage follows *DAL19* δ , and a minority of short-read coverage arrives at *DAL19* γ . This uneven split of *k*-mer coverage between the δ and γ regions together with the even split of *k*-mer coverage between the α and β regions is inconsistent with the expression of less than three *DAL19* transcript isoforms.

We attempted to assemble these transcript isoforms from the short RNA-seq reads using Trinity and Oases. However, the proportion of known *DAL19* transcript isoforms that were assembled with these methods was poor. For each *DAL19* transcript isoform and each assembler, we calculated a similarity score. This score consisted of the number of base matches of

the assembled transcript with the longest match to a *DAL19* transcript isoform divided by the number of bases in that *DAL19* transcript isoform. Trinity only assembled *DAL19*_{αψγ} and *DAL19*_{αψδ} with more than 90% similarity, and Oases only assembled *DAL19*_{βψγ} with more than 90% similarity (Table 2). Therefore, we developed a novel approach based on naive De Bruijn graph traversal coupled with kallisto (Bray et al., 2016) to reconstruct likely transcript isoforms containing 47-mers in the *DAL19* core region. Our method reconstructed *DAL19*_{αψγ}, *DAL19*_{αψδ}, and *DAL19*_{βψδ} at more than 90% similarity.

From the PacBio Iso-Seq data, we found at least one CCS read that was consistent over its entire length with only one of the transcripts *DAL19*_{αψγ}, *DAL19*_{αψδ}, *DAL19*_{βψγ}, or *DAL19*_{βψδ} (Figure 2). We also found CCS reads that mapped to the ψ region, but not to the α or β region. Although these reads could be evidence of *DAL19* transcripts without an α or β region, loss of 5′ ends due to the PacBio Iso-Seq library prep has been observed previously (Sharon et al., 2013; Gordon et al., 2015), and at least one read

contained *k*-mers that spanned the junction of *DAL19*_α and *DAL19*_ψ.

The Different *DAL19* Isoforms Have Distinct Expression Profiles in Early Bud Meristems

From previous experiments, we know that *DAL19* activity is upregulated in both needles and apical buds of cone-initiating *acrocona* shoots (Uddenberg et al., 2013). A reexamination using transcript-specific primers that amplify at similar efficiency indicates that it is the *DAL19*_{αψδ} isoform that dominates in cone-setting *acrocona* shoots (Supplementary Table S5).

To substantiate these findings and to study if the different *DAL19* isoforms are also differentially regulated in wild-type *Picea abies*, we analyzed their expression in buds of different identities (i.e., vegetative, male and female identity) using quantitative Real-Time PCR (qRT-PCR) (Figure 3). Templates for the qRT-PCR experiments were vegetative, female, and male buds collected during early bud development when most of the bud only consists of a large shoot-apical meristem (Figure 3A). As control samples, we included bud samples from a later phase of lateral organ formation (Figure 3B). *DAL19* isoform-specific primers with similar amplification efficiencies were used to assay relative transcript abundance in the different samples (Supplementary Figure S4). With the exception of transcripts with 5′ region *DAL19*_α in vegetative and male samples, all transcripts amplified at significantly higher levels in the early-phase meristems relative to buds collected at the later lateral-organ-formation-phase (Figures 3C–F). Relative to vegetative meristems, the isoforms containing the 5′ region *DAL19*_β amplified at a significantly higher level in male meristems (Figure 3D), whereas the isoforms containing the 5′ region *DAL19*_α amplified at significantly lower levels in female buds (Figure 3C). The isoforms containing the 3′ region *DAL19*_δ were significantly down-regulated in male meristems compared to vegetative meristems, whereas the isoforms containing the 3′ region *DAL19*_γ were significantly up-regulated in both female and male meristems (Figures 3E, F). Hence, by comparing ΔΔCT values in male and female buds relative to vegetative meristems, isoforms with *DAL19*_β and *DAL19*_γ appear to be upregulated in male buds, possibly reflecting an up-regulation of the *DAL19*_{βψγ} transcript in male meristems.

To test if the differences found in the qRT-PCR experiments are reflected in a spatial distribution of the *DAL19* isoforms in the early bud meristems, we conducted mRNA *in situ* hybridization experiments using isoform specific LNA probes hybridized against longitudinal sections of female, vegetative and male buds. Using a *DAL19*_δ specific probe we detected distinct hybridization signals in the epidermal cell layer of female and vegetative buds, and emerging lateral organs (Figures 4A–C). In male buds at a similar stage, the hybridization signal was less distinct in the epidermal layer and more evenly distributed into the underlying cell layers in the bud meristem (Figure 4D).

TABLE 2 | Assembled transcripts with best match with known full-length *DAL19* transcript isoforms.

<i>DAL19</i> transcript	Length of <i>DAL19</i> transcript	Assembly	Similarity	CIGAR
<i>DAL19</i> _{αψδ}	842	Oases	0.77	103S305M1D347M405S
<i>DAL19</i> _{αψγ}	813	Oases	0.87	103S13M1I291M1D241M105I267M139S
<i>DAL19</i> _{βψδ}	1095	Oases	0.80	872M405S
<i>DAL19</i> _{βψγ}	1067	Oases	0.97	1033M139S
<i>DAL19</i> _{αψδ}	842	Trinity	0.92	267S774M62S
<i>DAL19</i> _{αψγ}	813	Trinity	0.92	267S746M142S
<i>DAL19</i> _{βψδ}	1095	Trinity	0.68	274S767M62S
<i>DAL19</i> _{βψγ}	1067	Trinity	0.39	7M1I1M1I23M2I12M4D234M2I3M1D7M1D16M1D2M1I209M310S
<i>DAL19</i> _{αψδ}	842	cortexpy/kallisto	1.00	110S842M52S
<i>DAL19</i> _{αψγ}	813	cortexpy/kallisto	1.00	110S13M1I800M139S
<i>DAL19</i> _{βψδ}	1095	cortexpy/kallisto	0.92	1008M52S
<i>DAL19</i> _{βψγ}	1067	cortexpy/kallisto	0.45	177S31M6D41M2D4M1D4M5I2M4I3M1I8M2I10M3I5M3I2M1I35M2I8M2D1M2D472M524S
<i>DAL19</i> _{αψδ}	842	Nine bud samples	1.00	81S842M52S
<i>DAL19</i> _{αψγ}	813	Nine bud samples	1.00	81S13M1I800M69S
<i>DAL19</i> _{βψδ}	1095	Nine bud samples	0.96	1050M12S
<i>DAL19</i> _{βψγ}	1067	Nine bud samples	0.86	920M

For each *DAL19* transcript and each assembly the shortest assembled transcript with the longest sequence match to the *DAL19* transcript is reported. **Assembly:** name of assembly tool that generated the transcripts (see Assembly, Evaluation, and Visualization of *DAL19* Transcripts). **Similarity:** number of base matches to the *DAL19* transcript divided by the number of bases in the *DAL19* transcript. **CIGAR:** CIGAR string of transcript alignment to the *DAL19* transcript in SAMv1 format.

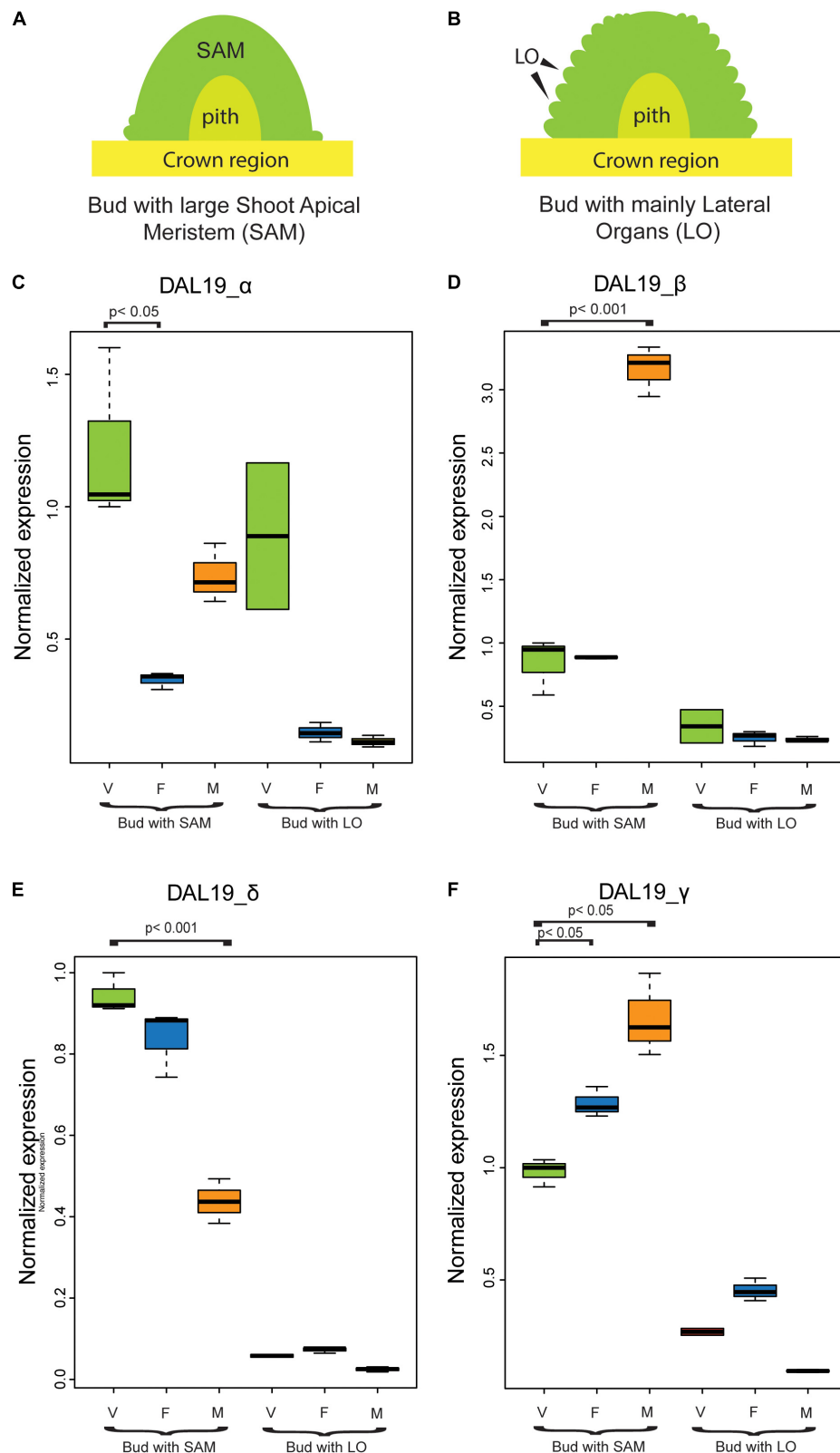


FIGURE 3 | Normalized expression of *DAL19* isoforms assayed by qRT-PCR. **(A)** Schematic representation of a bud with a large Shoot Apical Meristem (SAM) commonly found during late July or early August. **(B)** Schematic representation of a bud in late August or early September, when lateral organ formation is ongoing. **(C–F)** Normalized expression of the *DAL19* isoforms with α **(C)**, β **(D)**, δ **(E)**, and γ **(F)** exons. V, vegetative; ♀, female; ♂, male.

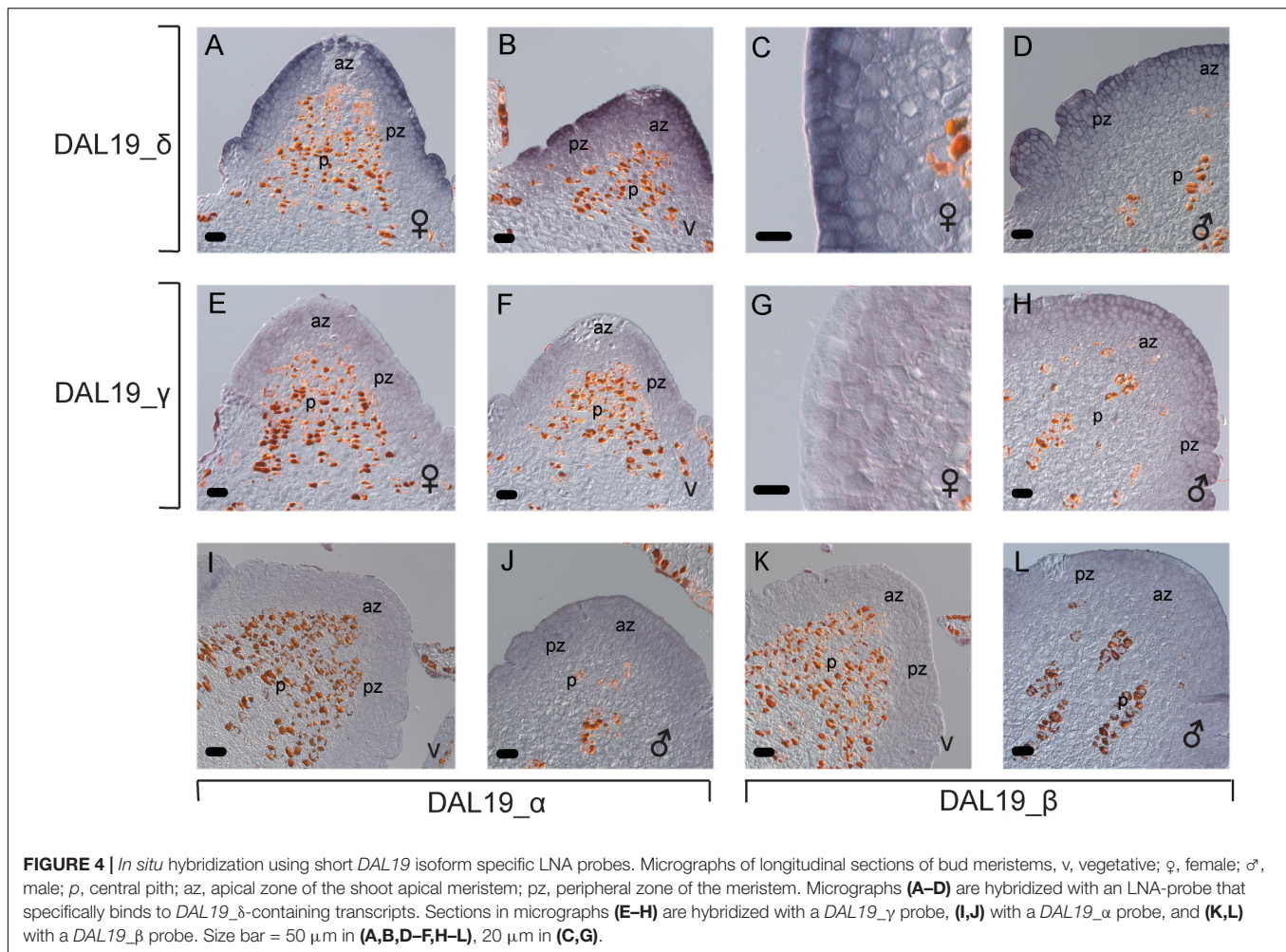


FIGURE 4 | *In situ* hybridization using short *DAL19* isoform specific LNA probes. Micrographs of longitudinal sections of bud meristems, v, vegetative; ♀, female; ♂, male; p, central pith; az, apical zone of the shoot apical meristem; pz, peripheral zone of the meristem. Micrographs (A–D) are hybridized with an LNA-probe that specifically binds to *DAL19_δ*-containing transcripts. Sections in micrographs (E–H) are hybridized with a *DAL19_γ* probe, (I,J) with a *DAL19_α* probe, and (K,L) with a *DAL19_β* probe. Size bar = 50 μ m in (A,B,D–F,H–L), 20 μ m in (C,G).

We detected a weak hybridization signal, almost complementary to that of *DAL19_δ*, in sections of female and vegetative buds hybridized with *DAL19_γ*-specific LNA-probes (Figures 4E–G). In longitudinal sections of female and vegetative buds, *DAL19_γ* hybridization was reduced or absent in the epidermal cell layer, whereas the signal was stronger in the underlying cell layers of the bud meristems (Figure 4G). In male buds, the signal from *DAL19_γ* LNA-probes matched that of *DAL19_δ* (Figure 4H). LNA-probes directed toward the 5' isoforms containing *DAL19_α* or *DAL19_β* showed a considerably weaker signal that is difficult to distinguish from the background (Figures 4I,J). Reliable signal of *DAL19_β* was only detected in male meristems, in a pattern that matched those of *DAL19_δ* and *DAL19_γ* (Figures 4D,H,L). *DAL19_α* signals were evenly distributed throughout the whole meristematic region of the bud but were lower in the central pith (Figure 4). LNA probes directed toward *Histone H2A*, included as positive experimental control, gave a characteristic patchy signal in dividing cells (Supplementary Figure S5). Taken together, this demonstrates that distinct *DAL19* isoforms are active in bud meristems of different identity and that cell-type-specific distribution of at least

the isoforms containing δ or γ can occur within one bud meristem.

Our Novel Assembly Approach Identifies 1,084 Putative MADS-Box Transcripts From Short-Read Sequencing

To assess if other novel MADS-box transcripts exist that are similar to the transcripts we discovered in *DAL19*, we applied our assembly method to RNA-seq data from nine meristematic bud samples to reconstruct likely transcript isoforms (see Assembly and Filtering of MADS-Box Transcripts in Section “Materials and Methods”). We merged assembled transcripts from all bud samples, annotated them with predicted protein domains, and mapped them back to MADS-box core regions. This resulted in 1,084 unique assembled transcripts, of which 1,073 could be annotated with at least one predicted protein domain, and of which 1,040 further mapped to a MADS-box core region. We manually curated these 1,040 assembled transcripts and performed multiple sequence alignment to arrive at 933 likely transcript isoforms. We present single nucleotide polymorphisms (SNPs), indels, and usage of alternate exons in Table 3.

TABLE 3 | Summary of transcripts aligned to MADS-core regions.

Core name	Associated MADS_MA	Tot. Number of transcripts	InDels	Alternate 5' exon	Alternate 3' exon	Short transcripts
DAL1	MA_453949	86	Yes	No	Yes, MA_369170 or MA_128164	Yes_only_MADS
DAL3	MA_141872, MA_10256834,	55	Yes	Yes	No	Yes_only_MADS due to frame shift
DAL3_like	MA_19007, MA_32676	33				
DAL4	MA_35712, MA_101463	139	No	Yes	No	Yes-3' frame shift
DAL10	MA_15122	84	Yes	No	No	Yes-5' frame shift
DAL11	MA_18048	22	Yes	No	No	Yes-3' frame shift
DAL12	MA_13933	28	Yes	No	No	Yes-3' frame shift
DAL13	MA_130755	25	Yes	No	No	No
DAL13_Like_a	MA_2191711	4	Yes	No	No	No
DAL13_Like_b	MA_8198323	3	Yes	No	Yes	Yes_only_MADS
DAL19	MA_16120, MA_162822, MA_329880,	90	Yes	Yes	Yes	Yes_only_K-domain
DAL21	MA_126898	21	Yes	No	No	Yes- 3' frame shift
DAL19_like_a/DAL31	MA_7272891	22	No	No	No	No
DAL32	MA_7858089, MA_138440	59	Yes	Yes	Yes, MA_G76145 or MA_109338	Yes_MA_576145
DAL33	MA_10079394, MA_276701	123	Yes	Yes	No	Yes-only MADS
DAL34						
DAL35	MA_9284799	20	No	No	No	No
DAL37	MA_18942	20	No	No	No	No
DAL38	MA_20180	12	No	No	Yes, MA_88284 or MA-63649	
DAL39	MA_94049	25	No	No	No	No
DAL40/JTL	MA_10432602	14	No	No	No	No
DAL41	MA_57186	26	No	No	No	No
DAL5	MA_32019	3	No	No	No	No
DAL9	MA_6279308	19	No	No	No	No
Sum		933				

Core name: The name of the MADS-box core sequence used in the analysis. Associated MADS_MA: The identifier of the scaffold harboring the first exon of the assembled transcripts. Total Number of transcripts: Number of transcripts associated with a specific isoform. InDels: Indication if InDels occurs among the aligned transcripts. Alternate 5' exon: Indication if mutually exclusive 5' exons were detected. Alternate 3' exons: Indication if mutually exclusive 3' exons were detected. Short transcripts: Indication if the gene has an alternative splice site selection at the 3' or 5' end of exons, rendering short transcripts.

We identified multiple isoforms of *DAL19*, which is consistent with the results presented in Section “Mapping of *DAL19* Transcripts to Genomic Assemblies Provides Support That These Transcripts Are Isoforms.” We assembled transcripts that matched the cloned and confirmed *DAL19* transcripts $\alpha\psi\gamma$, $\alpha\psi\delta$, $\beta\psi\gamma$, and $\beta\psi\delta$ with a similarity of 0.999, 0.998, 0.96, and 0.86, respectively (Table 2). The CIGAR strings for the alignments of the assembled transcripts to these *DAL19* transcripts (Table 2) indicated that the assembled transcript for $\beta\psi\delta$ was missing 147 bases, and that the assembled transcripts for the other three *DAL19* transcripts contained fewer than 100 extra bases on the 5' or 3' end of the transcript. Furthermore, we detected one additional *DAL19* transcript. This transcript contained a MADS-box that mapped to sequence scaffold MA_162822 in the *P. abies* genome assembly (V1.0).

As for *DAL19*, we also identified the use of alternate 5' exons for the core sequences of *DAL3*, *DAL3_like*, *DAL4*, *DAL32*, and *DAL33*. These first exon sequences mapped to distinct scaffolds in the *P. abies* genome assembly (V1.0) (Table 3, column 2) and nucleotide searches against the NCBI Conserved Domain Database (SDD) indicated that these exons all encode MADS-domains. However, the transcript isoform of *DAL33* that mapped to the genomic scaffold MA_10079394 differed substantially on the amino acid level, albeit not on the nucleotide level, from other known MADS-domains and lacked most conserved MADS-domain signature motifs such as for example the IKRIENS, RQVT, and the KKYELS motifs (Supplementary Figure S6A).

Apart from for *DAL19*, we identified the use of alternate 3' exons for the core sequences of the gene *DAL1* (Supplementary Figure S6B). As with *DAL19*, the usage of an alternate 3' exon

resulted in a premature stop codon and loss of the C-terminal domain, which harbors the signature motif of AGL6-like genes (DCEPTLQIGY). We also identified transcripts with a premature stop codon, often occurring shortly after the first exon, in *DAL3*, *DAL4*, *DAL12*, *DAL21*, *DAL32*, *DAL33* and *DAL38*. In the case of *DAL13*, we assembled three transcripts with several SNPs that were distributed over the entire ORF. In BLAST searches against the *P. abies* genome V1.0 these three *DAL13*-related transcripts mapped to different sequence scaffolds indicating that these transcripts were transcribed from different genes.

To provide independent evidence for the occurrence of different MADS-box gene isoforms, we mapped CCS reads to the assembled transcript isoforms and aggregated them by MADS-box core sequence. The number of CCS reads that only mapped to transcripts associated with a single MADS-box core sequence are presented in **Table 1**. The CCS reads were consistent with assembled transcript isoforms *DAL1*, *DAL3*, *DAL3_like_a*, *DAL4*, *DAL9*, *DAL19*, *DAL31*, *DAL32*, *DAL33*, *DAL35*, *DAL38*, *DAL37*, *DAL40/JTL*, and *DAL41* (**Table 1**).

174 Transcripts Are Differentially Expressed Across Three Bud Types

We assessed the differential expression of all assembled transcripts (1,084) across two male, three female, and three vegetative meristematic bud samples using a quasi-likelihood *F*-test (McCarthy et al., 2012) implemented in the edgeR package (Robinson et al., 2010). Hundred and seventy four transcripts were differentially expressed across all three groups at an FDR of 0.1 (**Supplementary Table S6**). We had previously assigned 152 of the 174 differentially expressed transcripts to 28 putative MADS-box genes (**Table 1**). The remaining ten putative MADS-box genes did not have any differentially expressed transcripts assigned. Of the 16 putative MADS-box genes for which there was evidence of expression from CCS reads, 14 genes had differentially expressed transcripts assigned.

All MADS-Box Genes Using Alternate First Exons Fall Into a Common Clade in Phylogenetic Analysis

To examine to what extent the usage of alternate exons in different MADS-box gene isoforms influences the position of MADS-box gene isoforms in the phylogenetic tree, maximum parsimony analyses were performed. In a first analysis, the nucleotide sequence spanning the open reading frame (ORF) of transcripts representing the identified MADS-box gene isoforms were aligned to MADS-box genes from *Picea abies*, *Pinus radiata*, *Arabidopsis thaliana*, and *Lycopersicon esculentum*, **Supplementary Figure S7A**. In a second analysis, only the MADS-box was used to determine the phylogenetic relationship (**Supplementary Figure S7B**).

In both analyses, and in agreement with previous published analyses, functionally related angiosperm MADS-box genes form monophyletic clades that often also have a gymnosperm sister-clade. For instance, *DAL2* from *Picea abies* and PrMADS1 grouped close to the AGAMOUS-clade as reported by (Tandre et al., 1995) and *DAL11-13* grouped with the angiosperm clade

harboring AP3 and PI as reported by (Sundstrom et al., 1999). In the first analysis, all *Picea abies* genes harboring alternate MADS-box sequences, e.g., *DAL3*, *DAL4*, *DAL19*, *DAL32*, and *DAL33*, fell into a common clade that formed a sister clade to the angiosperm TM3 clade. As expected, transcript isoforms belonging to the same MADS-box core sequence grouped together. In the second analysis, which was based on only the nucleotide sequences encoding MADS-domains, the majority of the isoforms still grouped in the same sub-clade. However, the internal relationships between isoforms and other MADS-box genes changed, and transcript isoforms assigned to the same core region were split up in different sub-clades. For example, in the first analysis, *DAL3* transcripts number 160 and 432 formed a well-supported subclade (bootstrap value 99) together with the transcript for *DAL3* deposited in Genbank (XY6654356). In the second analysis the *DAL3* sub-clade was split into two distinct sub-clades of which one grouped together with the transcript *DAL3 Like 966*. In addition, there was lack of support for the position of the *DAL33* transcript number 162 in the larger sub-clade of gymnosperm genes that formed a sister-clade to the TM3-subclade.

Apart from *DAL19*, assembled isoforms that used mutually exclusive exons in the 3' region were detected for the gene *DAL1*. In the phylogenetic reconstruction, *DAL1* and its corresponding isoforms did not group in the same sub-clade as the other genes with reported 5' isoforms. *DAL1*, *DAL14*, and the isoforms represented by transcripts number 580 and 802 instead formed a sister-clade to angiosperm AGL6/AGL2-like genes.

DISCUSSION

A multi-exon gene may be spliced into numerous variants through exon skipping, exon mutual exclusion, intron retention, or alternative splice site selection at the 3' or 5' end of exons (A3 or A5, and as reviewed by Reddy et al., 2013). Our analysis shows that all forms of alternative splicing occur in the MADS-box gene family in *P. abies*. Notably, the use of mutually exclusive first exons occurs strikingly often. We identified usage of mutually exclusive first exons in the genes *DAL3*, *DAL3_like*, *DAL4*, *DAL19*, *DAL32*, and *DAL33*. In these genes, the first exon encodes the DNA-binding MADS-domain. It is the MADS-domain that is responsible for the DNA-binding properties of the proteins and it has been demonstrated that this domain interacts with *cis*-regulatory DNA elements called CarG-boxes (Riechmann et al., 1996). Recently, it has also been demonstrated that small changes in the MADS-box amino acid composition might influence the affinity of the MADS-box to different CarG-box sequences (Smaczniak et al., 2017). The MADS-box domain of isoforms containing *DAL19_α* or *DAL19_β* differs in several amino acids: Two of the amino acids are in the highly conserved IKRIENS-motif and five amino acids differ in the region of the MADS-domain, which, according to structural models, is thought to encode β-sheets (**Supplementary Figure S6A**). A similar frequency of amino acid changes is also found in the MADS-domains of *DAL3*, *DAL3_like*, *DAL4*, *DAL19* and *DAL32* (**Supplementary Figure S6A**). We hypothesize that the

isoforms of these genes have different affinity to different CARG-box sequences and may thereby regulate different sets of target genes. In the gene *DAL33*, one of the isoforms has diverged considerably and has accumulated substitutions primarily in the first and second positions of the triplet codons that encode the MADS-domain signature motifs, indicating that this isoform of *DAL33* may confer a change or complete loss of DNA binding properties to the *DAL33* protein.

Apart from changes in DNA-binding properties, usage of mutually exclusive first exons may also confer different transcriptional activity. In fact, (Li et al., 2007) argue that mutually exclusive usage of first exons may constitute a distinct form of mutual exon exclusion because it also implies that the isoforms are transcribed from different promoters. In line with this, we detected differential expression of the isoforms containing *DAL19_α* or *DAL19_β* across female, vegetative, and male bud samples.

Several lines of evidence suggest that MADS-domain transcription factors form homo or hetero dimers, reviewed by Gramzow and Theissen (2010). In addition, both angiosperm and gymnosperm MADS-domain transcription factors active during reproductive development form multimeric complexes that have the ability to bind several CARG-boxes through DNA-looping (Kaufmann et al., 2005). Hence, the activity of a MADS-domain protein is determined both by its DNA-binding properties and its ability to interact with other MADS-domain proteins and associated proteins. Structural characterization of the intervening (I) and keratin-like (K) domains of SEPPALATA3 (SEP3) from *Arabidopsis thaliana* has demonstrated that the domains form two amphipathic alpha helices and that regularly spaced hydrophobic residues in those two helices are important for dimerization and for the formation of higher order tetramer complexes (Puranik et al., 2014). In the *DAL19* protein, the K-domain and the intervening region corresponds to the core region (ψ) defined here. We found no evidence of usage of mutually exclusive exons in the K-domain of the *P. abies* MADS-domain proteins although occasional retention of introns and alternative splice site selection could be observed. This suggests that alternative splicing does not primarily affect protein dimerization properties. This has implications for the interpretation of functional relevance of the short mature mRNA transcripts that lack DNA-binding MADS-domain i.e., *DAL19_ψα* or *DAL19_ψβ*. It is possible that transcription of a short transcript that has the ability to interact with other MADS-domain proteins but lacks the DNA-binding domain, may in fact act as a dominant negative protein.

We also identified use of mutually exclusive exons in the 3' region of the genes *DAL1*, and *DAL19*. In both genes, usage of an alternate 3' exon leads to a shorter mature mRNA transcript that lacks the C-terminal signature motifs (**Supplementary Figure S6B**). *DAL19_δ* harbors the signature motif (EVETQL) commonly found in TM3-like genes, whereas transcripts ending with the *DAL19_γ* exon yield a protein with a premature stop codon. Similarly, transcripts ending with *DAL1_α* harbor the AGL6-like motif (DCEPTLQIGY) whereas the usage of an alternate 3' C-terminal exon in *DAL1_β* leads to a pre-mature stop codon directly after the K-domain. It has been demonstrated

that the C-terminal of MADS-box genes are critical for functional specificity (Lamb and Irish, 2003) as the C-terminal may harbor activation domains or allow interactions with specific proteins (Litt and Irish, 2003). As judged by sequence comparison with the structurally characterized SEP3-protein (Puranik et al., 2014), the long protein isoforms of *DAL1* and *DAL19* have retained conserved hydrophobic residues in the last part of the K-domain that are important for dimerization and tetramerization. The short versions have retained all residues of importance for dimerization, but due to the usage of alternate 3' exons they lack part of the hydrophobic residues that have been shown to be of importance for tetramerization in SEP3. Hence, provided that the short transcript isoforms of *DAL1* and *DAL19* are translated into proteins, the resulting proteins may have retained their ability to dimerize but could have lost their ability to form higher order complexes and may in fact work as dominant negative proteins.

The occurrence of alternatively spliced *DAL19* and *DAL1* isoforms with premature stop codons is analogous to the MADS-box gene *FLOWERING LOCUS M (FLM)* in *A. thaliana*, which in its active form acts as a repressor of flower development (Pose et al., 2013). An increase in ambient temperature leads to alternative splicing of the *FLM* transcripts and the formation of a premature stop, which in turn triggers nonsense mediated decay (Sureshkumar et al., 2016; Capovilla et al., 2017). In this case, alternative splicing influences the amount of active protein in a specific cell or tissue. In our data, isoforms containing *DAL19_δ* were expressed at high levels in the epidermal layer of vegetative and female bud meristems, whereas the isoforms containing *DAL19_γ* showed a complementary expression pattern in the same meristem. This is strong evidence that cell-specific splicing may occur within a single meristem. Alternative splicing may down-regulate the amount of active protein in a specific cell through nonsense mediated decay or expression of dominant negative forms of the protein. This may contribute to the establishment of sharp boundaries within a meristem between cells that express the active form of the protein and cells that express the inactive or dominant negative form of the protein.

Furthermore, apart from use of mutually exclusive first or last exons, we also observe use of alternative 5' and 3' splice site selection in *DAL3*, *DAL4*, *DAL12*, *DAL21*, *DAL32*, *DAL4*, *DAL3*, *DAL33*, and *DAL38*. Studies in the model plant *Arabidopsis thaliana*, and other angiosperm species, indicate that this form of alternative splicing is more prevalent than the use of mutually exclusive exons (Severing et al., 2012; Zhang et al., 2015; Luo et al., 2017; Verhage et al., 2017). Among the *P. abies* MADS-box genes, these alternative splice sites often result in frame shifts and premature stop-codons shortly after the MADS-box region. This suggests that several MADS-box genes may be translated into both full-length proteins and micro-proteins.

Taken together, use of mutually exclusive first exons provides a means to express MADS-box genes from different promoters in a tissue or bud specific manner. The occurrence of different amino acids in the MADS-domains may confer varying DNA-binding properties to the resulting MADS proteins. This may affect the selection of down-stream target genes and may thereby change the regulation of bud development. The use of mutually exclusive last exons or alternative splice site selection provides a means to

either produce proteins with different function in a cell-specific manner, or to establish sharp boundaries between an active and an inactive isoform within a single meristem.

We also present a novel approach to transcriptome assembly. We used this approach to assemble *DAL19* transcript isoforms that match the cloned and confirmed *DAL19* isoforms $\alpha\psi\gamma$, $\alpha\psi\delta$, and $\beta\psi\gamma$, except for the 5' ends of the α and β regions and the 3' ends of the δ and γ regions. However, where the assembled and cloned transcripts diverged, CCS reads supported the assembled and not the cloned transcripts (dark green, dark blue, and dark red colors in **Supplementary Figures 2A,B**). It is likely that the cloned transcripts were shorter at the 3' ends because the 5'/3' RACE approach used to clone these transcripts is not guaranteed to clone the full length of the 3' transcript end. The divergence on the 5' end of the transcripts may be a rare variant in the *P. abies* reference.

We furthermore used our assembly approach on nine other bud samples to show that *P. abies* expresses hundreds of transcript isoforms containing one of 38 MADS-box core sequences. We found 933 plausible transcripts of which 152 were differentially expressed across bud types, and of which the majority clustered in an expected manner with known DAL transcripts in phylogenetic analyses of full-length transcripts and of only the MADS-box region. A minority of plausible MADS-box transcripts displayed different clustering between the two phylogenetic analyses. This could be due to assembly errors or it could reflect real gene fusions in the evolutionary history of these transcripts.

Our assembly approach appears to have high sensitivity to transcript isoforms of the MADS-box gene family in *P. abies*. However, further work is needed to establish the false positive rate of our method, how this method performs more generally for the *P. abies* transcriptome, and how it can be used for assembling the transcriptomes of other organisms. Our method generates a candidate transcript for every possible 5' to 3' path through a De Bruijn graph representation of RNA-seq reads, and it relies on kallisto to filter these candidate transcripts down to a reasonable number. Unfortunately, the number of candidate transcripts generated scales exponentially with the number of branches between any incoming and outgoing tip in the De Bruijn graph. Therefore, genes with more splice variants or sequence polymorphisms could cause the number of candidate transcripts to grow to a size that kallisto cannot manage. Our approach does not handle cycles in the graph, although links as described by Turner et al. (2018) could be used to traverse small cycles. Finally, our approach cannot detect truncated transcripts together with full-length transcripts, as candidate transcripts may only start on incoming tips of the De Bruijn graph. However, allowing candidate transcripts to also start from *k*-mers with a large coverage increase compared to their incoming neighbor *k*-mers might allow the detection of truncated transcripts.

Phylogenetic reconstructions of the MADS-box gene family have shown that functionally related genes group together in monophyletic clades (see e.g., Becker and Theissen, 2003 and references within). Based on genome-wide analyses and transcriptome data it has also been suggested that in gymnosperms, MADS-box genes orthologous to *DAL19* have

undergone a series of duplication events leading to a rapid expansion in the number of genes and the formation of a *DAL19*-clade in several gymnosperm lineages (Gramzow et al., 2014; Chen et al., 2017).

Based on the phylogenetic reconstruction of the MADS-box gene family and their transcript isoforms, all *P. abies* MADS-box genes that display alternative splicing of mutually exclusive first exons grouped together in the *DAL19*-clade. Hence, the observed complexity in this clade may not only be due to duplication events but also due to alternative splicing and usage of mutually exclusive first exons. We also note that phylogenies based solely on the conserved MADS-box may lead to an overestimation of the number of genes and changes in the tree topology, which in turn could influence the interpretation of the phylogenetic relationships between different MADS-box genes. Functional characterization of the angiosperm genes within the *TM3-like* clade has demonstrated that several genes are involved in the transition from vegetative to reproductive growth. We hypothesize that increased complexity in terms of number of genes and usage of transcript isoforms in the gymnosperm sister-clade to *TM3-like* genes reflects a complex genetic regulation of vegetative to reproductive phase change and cone-setting in conifers and other gymnosperms. This regulation possibly involves responses to environmental factors such as ambient temperature or daylight that may influence splicing and the transcription of different transcript isoforms. Temperature is, in fact, used as a predictor of cone initiation in *P. abies* (Lindgren et al., 1977), and we hypothesize that alternative splicing is one of the molecular mechanisms employed by the tree to determine whether or not to produce cones.

AUTHOR CONTRIBUTIONS

SA collected the plant material, performed qPCR, and discovered putative MADS-box core sequences. SA and VN prepared the RNA and performed the cloning. ON, OE, and JS provided funding for the Illumina sequencing. ON provided funding for the PacBio sequencing. ND, VN, and NS pre-processed the short-read and long-read sequencing data. WK developed the assembly method and performed the assembly, filtering, mapping, and differential expression analysis of transcripts. JS curated the transcript isoforms and performed the phylogenetic analysis. SA and JS performed *in situ* hybridization. JS supervised SA and OE supervised WK. JS and ON supervised VN. SA, WK, and JS wrote substantial portions of the manuscript. SA, WK, and JS designed the study. All authors edited the manuscript.

FUNDING

This work was supported by the Knut and Alice Wallenberg Foundation and the Swedish Governmental Agency for Innovation Systems. NS was supported by Trees and Crop for the Future (TC4F) project. SA was supported by Formas Grant Dnr 239-2013-650 and VN by the SLU Plant Breeding platform.

ACKNOWLEDGMENTS

We thank the Swedish National Genomics Infrastructure hosted at SciLifeLab, the National Bioinformatics Infrastructure Sweden (NBIS) for providing computational assistance, and the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) for providing computational infrastructure. We thank Daniel Uddenberg, Ewa Mellerowicz, Julia Haas, and Soile Jokipii-Lukkari for their contributions of samples for PacBio Iso-Seq sequencing. We thank Kiran V Garimella for his code contributions to cortexpy, and for suggesting and writing a proof of concept that an approach based on Cortex graphs could be used to assemble long transcripts of

DAL19. We thank Anders F. Andersson for helpful feedback and ideas during the development of our transcript assembly method.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.01625/full#supplementary-material>

All sequencing data associated with this paper can be obtained from the European Nucleotide Archive with the accession numbers PRJEB27247 and ERP109307.

REFERENCES

- Abdel-Ghany, S. E., Hamilton, M., Jacobi, J. L., Ngam, P., Devitt, N., Schilkey, F., et al. (2016). A survey of the sorghum transcriptome using single-molecule long reads. *Nat. Commun.* 7:11706. doi: 10.1038/ncomms11706
- Azevedo, H., Lino-Neto, T., and Tavares, R. M. (2003). An improved method for high-quality RNA isolation from needles of adult maritime pine trees. *Plant Mol. Biol. Rep.* 21, 333–338. doi: 10.1007/BF02772582
- Becker, A., and Theissen, G. (2003). The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Mol. Phylogenet. Evol.* 29, 464–489. doi: 10.1016/S1055-7903(03)00207-0
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57, 289–300.
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 34, 525–527. doi: 10.1038/nbt.3519
- Capovilla, G., Symeonidi, E., Wu, R., and Schmid, M. (2017). Contribution of major FLM isoforms to temperature-dependent flowering in *Arabidopsis thaliana*. *J. Exp. Bot.* 68, 5117–5127. doi: 10.1093/jxb/erx328
- Carlsbecker, A., Sundstrom, J. F., Englund, M., Uddenberg, D., Izquierdo, L., Kvarnheden, A., et al. (2013). Molecular control of normal and acrocona mutant seed cone development in Norway spruce (*Picea abies*) and the evolution of conifer ovule-bearing organs. *New Phytol.* 200, 261–275. doi: 10.1111/nph.12360
- Chen, F., Zhang, X., Liu, X., and Zhang, L. (2017). Evolutionary analysis of MIKC^c-type MADS-box genes in gymnosperms and angiosperms. *Front. Plant Sci.* 8:895. doi: 10.3389/fpls.2017.00895
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., et al. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biol.* 17:13. doi: 10.1186/s13059-016-0881-8
- Dorca-Fornell, C., Gregis, V., Grandi, V., Coupland, G., Colombo, L., and Kater, M. M. (2011). The Arabidopsis SOC1-like genes AGL42, AGL71 and AGL72 promote flowering in the shoot apical and axillary meristems. *Plant J.* 67, 1006–1017. doi: 10.1111/j.1365-3113X.2011.04653.x
- Dwyer, T., Koren, Y., and Marriott, K. (2006). IPSEP-COLA: an incremental procedure for separation constraint layout of graphs. *IEEE Trans. Vis. Comput. Graph* 12, 821–828. doi: 10.1109/TVCG.2006.156
- Gordon, S. P., Tseng, E., Salamov, A., Zhang, J., Meng, X., Zhao, Z., et al. (2015). Widespread polycistronic transcripts in fungi revealed by single-molecule mRNA sequencing. *PLoS One* 10:e0132628. doi: 10.1371/journal.pone.0132628
- 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, 644–652. doi: 10.1038/nbt.1883
- Gramzow, L., and Theissen, G. (2010). A hitchhiker's guide to the MADS world of plants. *Genome Biol.* 11:214. doi: 10.1186/gb-2010-11-6-214
- Gramzow, L., Weilandt, L., and Theissen, G. (2014). MADS goes genomic in conifers: towards determining the ancestral set of MADS-box genes in seed plants. *Ann. Bot.* 114, 1407–1429. doi: 10.1093/aob/mcu066
- Hannerz, M., and Sundström, J. F. (2006). "I är blommar granen", in: *PlantAktuellt*, ed. E. Valinger (Uppsala: Skogforsk).
- Hoang, N. V., Furtado, A., Mason, P. J., Marquardt, A., Kasirajan, L., Thiruganasambandam, P. P., et al. (2017). A survey of the complex transcriptome from the highly polyploid sugarcane genome using full-length isoform sequencing and de novo assembly from short read sequencing. *BMC Genomics* 18:395. doi: 10.1186/s12864-017-3757-8
- Jasaitiene, D., Valiukeviciene, S., Linkeviciute, G., Rausitis, R., Jasiuniene, E., and Kazys, R. (2011). Principles of high-frequency ultrasonography for investigation of skin pathology. *J. Eur. Acad. Dermatol. Venereol.* 25, 375–382. doi: 10.1111/j.1468-3083.2010.03837.x
- Karlgen, A., Carlsson, J., Gyllenstrand, N., Lagercrantz, U., and Sundstrom, J. F. (2009). Non-radioactive in situ hybridization protocol applicable for Norway spruce and a range of plant species. *J. Vis. Exp.* 26:e1205. doi: 10.3791/1205
- Kaufmann, K., Melzer, R., and Theissen, G. (2005). MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. *Gene* 347, 183–198. doi: 10.1016/j.gene.2004.12.014
- Kopylova, E., Noé, L., and Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* 28, 3211–3217. doi: 10.1093/bioinformatics/bts611
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109
- Lamb, R. S., and Irish, V. F. (2003). Functional divergence within the APETALA3/PISTILLATA floral homeotic gene lineages. *Proc. Natl. Acad. Sci. U.S.A.* 100, 6558–6563. doi: 10.1073/pnas.0631708100
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *Genomics* 2013, 1–3.
- Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. doi: 10.1093/bioinformatics/bty191
- Li, Q., Lee, J. A., and Black, D. L. (2007). Neuronal regulation of alternative pre-mRNA splicing. *Nat. Rev. Neurosci.* 8, 819–831. doi: 10.1038/nrn2237
- Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. doi: 10.1093/bioinformatics/btl158
- Lindgren, K., Ekberg, L., and Eriksson, G. (1977). External factors influencing female flowering in *Picea abies* (L.) Karst. *Stud. For. Suec.* 142, 1–53.
- Litt, A., and Irish, V. F. (2003). Duplication and diversification in the APETALA1/FRUITFULL floral homeotic gene lineage: implications for the evolution of floral development. *Genetics* 165, 821–833.
- Liu, X., Mei, W., Soltis, P. S., Soltis, D. E., and Barbazuk, W. B. (2017). Detecting alternatively spliced transcript isoforms from single-molecule long-read sequences without a reference genome. *Mol. Ecol. Res.* 17, 1243–1256. doi: 10.1111/1755-0998.12670
- Luo, X., Xu, L., Liang, D., Wang, Y., Zhang, W., Zhu, X., et al. (2017). Comparative transcriptomics uncovers alternative splicing and molecular

- marker development in radish (*Raphanus sativus* L.). *BMC Genomics* 18:505. doi: 10.1186/s12864-017-3874-4
- Ma, H., Yanofsky, M. F., and Meyerowitz, E. M. (1991). AGL1-AGL6, an Arabidopsis gene family with similarity to floral homeotic and transcription factor genes. *Genes Dev.* 5, 484–495. doi: 10.1101/gad.5.3.484
- Mandadi, K. K., and Scholthof, K. B. G. (2015). Genome-wide analysis of alternative splicing landscapes modulated during plant-virus interactions in *Brachypodium distachyon*. *Plant Cell* 27, 71–85. doi: 10.1105/tpc.114.133991
- Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., et al. (2015). CDD: NCBI's conserved domain database. *Nucleic Acids Res.* 43, D222–D226. doi: 10.1093/nar/gku1221
- Marquez, Y., Brown, J. W., Simpson, C., Barta, A., and Kalyna, M. (2012). Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis. *Genome Res.* 22, 1184–1195. doi: 10.1101/gr.13410.6.111
- McCarthy, D. J., Chen, Y., and Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* 40, 4288–4297. doi: 10.1093/nar/gks042
- Mortazavi, A., Williams, B. A., Mccue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628. doi: 10.1038/nmeth.1226
- Nystedt, B., Street, N. R., Wetterbom, A., Zuccolo, A., Lin, Y. C., Scofield, D. G., et al. (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature* 497, 579–584. doi: 10.1038/nature12211
- O'Maileidigh, D. S., Graciet, E., and Wellmer, F. (2014). Gene networks controlling *Arabidopsis thaliana* flower development. *New Phytol.* 201, 16–30. doi: 10.1111/nph.12444
- Pose, D., Verhage, L., Ott, F., Yant, L., Mathieu, J., Angenent, G. C., et al. (2013). Temperature-dependent regulation of flowering by antagonistic FLM variants. *Nature* 503, 414–417. doi: 10.1038/nature12633
- Puranik, S., Acajjaoui, S., Conn, S., Costa, L., Conn, V., Vial, A., et al. (2014). Structural basis for the oligomerization of the MADS domain transcription factor SEPALLATA3 in Arabidopsis. *Plant Cell* 26, 3603–3615. doi: 10.1105/tpc.114.127910
- Reddy, A. S., Marquez, Y., Kalyna, M., and Barta, A. (2013). Complexity of the alternative splicing landscape in plants. *Plant Cell* 25, 3657–3683. doi: 10.1105/tpc.113.117523
- Riechmann, J. L., Wang, M., and Meyerowitz, E. M. (1996). DNA-binding properties of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA and AGAMOUS. *Nucleic Acids Res.* 24, 3134–3141. doi: 10.1093/nar/24.16.3134
- Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S. D., et al. (2010). De novo assembly and analysis of RNA-seq data. *Nat. Methods* 7, 909–912. doi: 10.1038/nmeth.1517
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616
- Schulz, M. H., Zerbino, D. R., Vingron, M., and Birney, E. (2012). Oases: Robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 28, 1086–1092. doi: 10.1093/bioinformatics/bts094
- Schwarz-Sommer, Z., Huijser, P., Nacken, W., Saedler, H., and Sommer, H. (1990). Genetic control of flower development by homeotic genes in *Antirrhinum majus*. *Science* 250, 931–936. doi: 10.1126/science.250.4983.931
- Severing, E. I., Van Dijk, A. D., Morabito, G., Busscher-Lange, J., Immink, R. G., and Van Ham, R. C. (2012). Predicting the impact of alternative splicing on plant MADS domain protein function. *PLoS One* 7:e30524. doi: 10.1371/journal.pone.0030524
- Sharon, D., Tilgner, H., Grubert, F., and Snyder, M. (2013). A single-molecule long-read survey of the human transcriptome. *Nat. Biotechnol.* 31, 1009–1014. doi: 10.1038/nbt.2705
- Smaczniak, C., Angenent, G. C., and Kaufmann, K. (2017). SELEX-Seq: a method to determine DNA binding specificities of plant transcription factors. *Methods Mol. Biol.* 1629, 67–82. doi: 10.1007/978-1-4939-7125-1_6
- Soneson, C., Love, M. I., and Robinson, M. D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res.* 4:1521. doi: 10.12688/f1000research.7563.1
- Steijger, T., Abril, J. F., Engström, P. G., Kokocinski, F., Abril, J. F., Akerman, M., et al. (2013). Assessment of transcript reconstruction methods for RNA-seq. *Nat. Methods* 10, 1177–1184. doi: 10.1038/nmeth.2714
- Sundell, D., Mannapperuma, C., Netotea, S., Delhomme, N., Lin, Y. C., and Sjdin, A. A. (2015). The plant genome integrative explorer resource: plantgenie.org. *New Phytol.* 208, 1149–1156. doi: 10.1111/nph.13557
- Sundstrom, J., Carlsbecker, A., Svensson, M. E., Svenson, M., Johanson, U., Theissen, G., et al. (1999). MADS-box genes active in developing pollen cones of Norway spruce (*Picea abies*) are homologous to the B-class floral homeotic genes in angiosperms. *Dev. Genet.* 25, 253–266. doi: 10.1002/(SICI)1520-6408(1999)25:3<253::AID-DVG8>3.0.CO;2-P
- Sureshkumar, S., Dent, C., Seleznev, A., Tasset, C., and Balasubramanian, S. (2016). Nonsense-mediated mRNA decay modulates FLM-dependent thermosensory flowering response in Arabidopsis. *Nat. Plants* 2:16055. doi: 10.1038/nplants.2016.55
- Swofford, D. L. (2002). *PAUP Phylogenetic Analysis Using Parsimony (and other methods)*. Sunderland, MA: Sinauer Associates.
- Tandre, K., Albert, V. A., Sundas, A., and Engstrom, P. (1995). Conifer homologues to genes that control floral development in angiosperms. *Plant Mol. Biol.* 27, 69–78. doi: 10.1007/BF00019179
- Thatcher, S. R., Zhou, W., Leonard, A., Wang, B. B., Beatty, M., Zastrow-Hayes, G., et al. (2014). Genome-wide analysis of alternative splicing in ze mays: landscape and genetic regulation. *Plant Cell* 26, 3472–3487. doi: 10.1105/tpc.114.130773
- Turner, I., Garimella, K. V., Iqbal, Z., and McVean, G. (2018). Integrating long-range connectivity information into de Bruijn graphs. *Bioinformatics* 34, 2556–2565. doi: 10.1093/bioinformatics/bty157
- Uddenberg, D., Reimegard, J., Clapham, D., Almqvist, C., Von Arnold, S., Emanuelsson, O., et al. (2013). Early cone setting in *Picea abies* acrocona is associated with increased transcriptional activity of a MADS box transcription factor. *Plant Physiol.* 161, 813–823. doi: 10.1104/pp.112.207746
- Verhage, L., Severing, E. I., Bucher, J., Lammers, M., Busscher-Lange, J., Bonnema, G., et al. (2017). Splicing-related genes are alternatively spliced upon changes in ambient temperatures in plants. *PLoS One* 12:e0172950. doi: 10.1371/journal.pone.0172950
- Wang, B., Tseng, E., Regulski, M., Clark, T. A., Hon, T., Jiao, Y., et al. (2016). Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nat. Commun.* 7:11708. doi: 10.1038/ncomms11708
- Winter, K. U., Becker, A., Munster, T., Kim, J. T., Saedler, H., and Theissen, G. (1999). MADS-box genes reveal that gnepophytes are more closely related to conifers than to flowering plants. *Proc. Natl. Acad. Sci. U.S.A.* 96, 7342–7347. doi: 10.1073/pnas.96.13.7342
- Xu, Z., Peters, R. J., Weirather, J., Luo, H., Liao, B., Zhang, X., et al. (2015). Full-length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of *Salvia miltiorrhiza* and tanshinone biosynthesis. *Plant J.* 82, 951–961. doi: 10.1111/tpj.12865
- Yeku, O., and Frohman, M. A. (2011). Rapid amplification of cDNA ends (RACE). *Methods Mol. Biol.* 703, 107–122. doi: 10.1007/978-1-59745-248-9_8
- Zhang, B., Liu, Z. X., Ma, J., Song, Y., and Chen, F. J. (2015). Alternative splicing of the AGAMOUS orthologous gene in double flower of *Magnolia stellata* (Magnoliaceae). *Plant Sci.* 241, 277–285. doi: 10.1016/j.plantsci.2015.10.017
- Zhang, J., Kobert, K., Flouri, T., and Stamatakis, A. (2014). PEAR: a fast and accurate illumina paired-end reAd mergeR. *Bioinformatics* 30, 614–620. doi: 10.1093/bioinformatics/btt593

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 © 2018 Akhter, Kretzschmar, Nordan, Delhomme, Street, Nilsson, Emanuelsson and Sundström. 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 Functional Change and Deletion of *FLC* Homologs Contribute to the Evolution of Rapid Flowering in *Boechnera stricta*

Cheng-Ruei Lee^{1,2,3*}, Jo-Wei Hsieh¹, M. E. Schranz⁴ and Thomas Mitchell-Olds⁵

¹ Institute of Ecology and Evolutionary Biology, National Taiwan University, Taipei, Taiwan, ² Institute of Plant Biology, National Taiwan University, Taipei, Taiwan, ³ Genome and Systems Biology Degree Program, National Taiwan University, Taipei, Taiwan, ⁴ Biosystematics Group, Wageningen University & Research, Wageningen, Netherlands, ⁵ Department of Biology, Duke University, Durham, NC, United States

OPEN ACCESS

Edited by:

Verónica S. Di Stilio,
University of Washington,
United States

Reviewed by:

Jill Christine Preston,
University of Vermont, United States
Kentaro K. Shimizu,
Universität Zürich, Switzerland

*Correspondence:

Cheng-Ruei Lee
chengrueilee@ntu.edu.tw

Specialty section:

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

Received: 20 April 2018

Accepted: 03 July 2018

Published: 31 July 2018

Citation:

Lee C-R, Hsieh J-W, Schranz ME and Mitchell-Olds T (2018) The Functional Change and Deletion of *FLC* Homologs Contribute to the Evolution of Rapid Flowering in *Boechnera stricta*. *Front. Plant Sci.* 9:1078. doi: 10.3389/fpls.2018.01078

Differences in the timing of vegetative-to-reproductive phase transition have evolved independently and repeatedly in different plant species. Due to their specific biological functions and positions in pathways, some genes are important targets of repeated evolution – independent mutations on these genes caused the evolution of similar phenotypes in distantly related organisms. While many studies have investigated these genes, it remains unclear how gene duplications influence repeated phenotypic evolution. Here we characterized the genetic architecture underlying a novel rapid-flowering phenotype in *Boechnera stricta* and investigated the candidate genes *BsFLC1* and *BsFLC2*. The expression patterns of *BsFLC1* suggested its function in flowering time suppression, and the deletion of *BsFLC1* is associated with rapid flowering and loss of vernalization requirement. In contrast, *BsFLC2* did not appear to be associated with flowering and had accumulated multiple amino acid substitutions in the relatively short evolutionary timeframe after gene duplication. These non-synonymous substitutions greatly changed the physicochemical properties of the original amino acids, concentrated non-randomly near a protein-interacting domain, and had greater substitution rate than synonymous changes. Here we suggested that, after recent gene duplication of the *FLC* gene, the evolution of rapid phenology was made possible by the change of *BsFLC2* expression pattern or protein sequences and the deletion of *BsFLC1*.

Keywords: *Boechnera stricta*, *FLOWERING LOCUS C (FLC)*, flowering time, gene duplication, repeated evolution

INTRODUCTION

Disentangling the genetic architecture underlying evolutionary changes is a critical task for evolutionary geneticists, and the evolution of parallel phenotypic changes is particularly intriguing. For example, did independent evolution of the same phenotypes occur through *de novo* genetic changes in the same pathway, the same gene, or even the same nucleotide? Martin and Orgogozo (2013) reviewed many cases of genetic changes underlying phenotypic differences and found more than 100 cases describing “genetic hotspots” under repeated evolution.

Repeated evolution through the same gene could happen in two ways: independent gain or loss of function in separate lineages. The latter should be more frequent than the former because independent gain-of-function events often require the same mutation on the same functional site. Independent loss-of-function events, on the other hand, can happen in many ways, such as point mutation creating a premature stop codon, large insertions disrupting a gene, deletions removing the whole gene or important functional sites, or gene disruption by structural rearrangements. Recent gene duplication adds another level of complexity to this question. After gene duplication, duplicated homologs might undergo neo-functionalization or sub-functionalization, which will facilitate the gain of new functions. For loss-of-function events to happen, many possibilities exist depending on how the genes were duplicated. For example, if duplicated copies retain their original function, such functional redundancy makes further loss-of-function evolution difficult. On the other hand, one mutation in the original copy can cause loss of function if the duplicated copy does not retain the original function. This can happen if (1) only a portion of the gene was duplicated, creating a truncated and non-functional new copy, (2) the duplication event did not cover important regulatory regions flanking the gene, making the new copy cease to express in the same tissue or time, or (3) the new copy was intact and had the same spatiotemporal patterns of expression but later underwent amino acid substitutions changing the protein function. Note that while possibility (1) and (2) indicate a genetic change caused by the duplication event itself, possibility (3) focuses on changes after the duplication.

The onset of flowering marks the transition from vegetative to reproductive phases of plants and is under strong natural selection (Anderson et al., 2011; Munguía-Rosas et al., 2011). In some Brassicaceae, the requirement for vernalization before flowering is polymorphic, with vernalization requirement as the ancestral state. In *Arabidopsis thaliana*, this is mainly controlled by the natural variation of two floral suppressor genes, *FRIGIDA* (*FRI*) and *FLOWERING LOCUS C* (*FLC*), whose loss- or decrease-of-function mutations cause rapid flowering without vernalization requirement (Michaels and Amasino, 1999; Gazzani et al., 2003; Michaels et al., 2003; Lempe et al., 2005; Shindo et al., 2005; Werner et al., 2005). In this species, this phenotype has independently evolved from different mutations in the same genes (Johanson et al., 2000; Le Corre et al., 2002; Gazzani et al., 2003; Michaels et al., 2003; Lempe et al., 2005; Shindo et al., 2005; Werner et al., 2005; Méndez-Vigo et al., 2011; Li et al., 2014), constituting a good example of repeated evolution through the same gene.

Here we investigate the evolution of rapid lifecycle in *Boechnera stricta* (Brassicaceae), a perennial wild relative of *Arabidopsis*. We ask whether parallel loss of vernalization responsiveness in a rapid-flowering phenotype of *B. stricta* involved mutations in one or both *FLC* homologs derived from a recent duplication event. We identified the candidate gene and its duplicated homolog, investigated their effects and expression patterns, and showed that the duplicated copy may have lost its flowering-related

functions, allowing the evolution of rapid life cycle through deletion of the original *FLC* copy.

MATERIALS AND METHODS

Plant Material and Flowering Time Estimation

Boechnera stricta is a short-lived and predominantly self-fertilizing perennial native to the Rocky Mountains in North America. It has two subspecies (Lee and Mitchell-Olds, 2011; Lee et al., 2017) differing in many life history traits, with the EAST subspecies flowering faster than the WEST subspecies under most greenhouse and vernalization conditions (Lee and Mitchell-Olds, 2013). Previously, a major QTL, *nFT*, containing the candidate flowering time gene *FT*, was identified controlling phenological traits and local fitness in various environments (Anderson et al., 2011, 2012, 2014). Further investigations showed that the two parental alleles are identical in coding sequences but different in expression (Lee et al., 2014). In this study we investigated natural variation in vernalization requirement, which has not been previously studied in this species.

From a sample of ~250 *B. stricta* accessions from the Northern Rocky Mountains (Lee and Mitchell-Olds, 2011), we identified one accession with very rapid flowering time from the EAST subspecies (family number 24B from the Moonrise Ridge population from Idaho, 44° 39' N, 114° 32' W, hereafter MR24). In the absence of vernalization in the greenhouse a typical *B. stricta* plant did not flower until 4–6 months, whereas the MR24 accession flowered within 2 months. We chose 14 accessions from the Moonrise Ridge population and performed complete randomized block experiments in the greenhouse to characterize flowering phenology with or without vernalization. To minimize maternal effects, we used seeds from accessions that were grown in the greenhouse for at least one generation. The greenhouse experiment consisted of 14 accessions, with seven from the faster and seven from the slower phenology groups. Each block contained 84 individuals with six plants from each accession, where three individuals were assigned to vernalization and three to the non-vernalization treatment. A total of seven blocks (588 plants) were used in this experiment. Like *A. thaliana*, *B. stricta* is a naturally self-fertilizing species, and sibling plants of the same accession were treated as replicated genetic clones throughout this study. Seeds were stratified in 4°C for 4 weeks, and seedlings were grown in “cone-tainer” racks (Stuewe & Sons Inc., Tangent, OR, United States) in the same environment (16-h days and 20°C ambient temperature) described in Lee and Mitchell-Olds (2013). When 2-month old, plants in the vernalization treatment were moved to 4°C, 10-h days for 6 weeks, and plants in the non-vernalization treatment remained in the greenhouse. The vernalization treatments for all following experiments were performed under the same condition (4°C, 10-h days) to simulate field environments during late winter and early spring. Since such treatment changed both temperature and day length, we recognize the caveat that we may not be able to separate effects from the two factors on flowering. We recorded flowering time (days to first flower) and leaf number

when flowering. For plants with vernalization, the duration under vernalization was subtracted from the estimation of flowering time, thus recording the number of days under warm conditions. The data were analyzed with mixed-effect ANOVA using days to first flowering or leaf number when first flowering as the response variable. Treatment (vernalization/non-vernalization) was used as fixed effect. Block, accession, and accession by treatment interaction were treated as random effects in JMP 8 (SAS, Cary, NC, United States).

Candidate Gene Profiling for Rapid Phenology

To identify candidate genes controlling the rapid phenology of MR24, we compared MR24 with a local accession (MR7A), whose flowering was significantly accelerated by vernalization. Using genomic DNA and cDNA from rosette leaves, we performed PCR and reverse transcription PCR (RT-PCR) on *FLC*, an important gene controlling the vernalization requirement in *Arabidopsis* (Michaels and Amasino, 1999; Sheldon et al., 2000; Shindo et al., 2005). All primer sequences are available in Supplementary Table S1. A previous study identified two *FLC* homologs in the *Boechera* genome (Schrantz et al., 2007). This gene duplication is not due to polyploidy or tandem duplication, but rather by a transposition duplication (Lee et al., 2017). *BsFLC1* locates in the ancestral syntenic position (Scaffold13175 on chromosome 6) with *Arabidopsis FLC*, while *BsFLC2* was duplicated and transposed to Scaffold18351 on *B. stricta* chromosome 5. Based on this information, we designed copy-specific PCR primers for *B. stricta FLC* RT-PCR (*BsFLC1*: primer FLC-B and FLC1-C, *BsFLC2*: primer FLC-B and FLC2-C, Supplementary Table S1). To further investigate whether the lack of *BsFLC1* PCR amplification is due to a missing *BsFLC1* or mutations in primer binding sites, we used three forward primers conserved between *BsFLC1* and *BsFLC2* (FLC-B, FLC-F, FLC-G, Supplementary Table S1) and three *BsFLC1*-specific reverse primers (FLC1-C, FLC1-D, FLC1-E, Supplementary Table S1) to generate nine PCR amplicons for the genomic DNA of *BsFLC1*.

We further used Nanopore long-read sequencing technology to identify the extent of deletion. Genomic DNA of the MR24 accession was extracted using a modified CTAB protocol (Doyle and Doyle, 1987), and the library was made with the Nanopore Rapid Sequencing Kit (SQK-RAD004) and sequenced in one R9.4 flow cell. All library preparation and sequencing steps were performed according to standard Nanopore protocols. Resulting reads in fastq format were mapped to the *B. stricta* reference genome v1.2 (Lee et al., 2017) using GraphMap (Sović et al., 2016) and visualized with Integrative Genomics Viewer (Robinson et al., 2011). We designed primer pairs spanning the inferred chromosomal deletion and confirmed the deletion with PCR and Sanger sequencing.

Test of *BsFLC1* Effects on Flowering Time

To test whether the lack of *BsFLC1* co-segregated with rapid phenology in progeny of accession MR24, we

performed a controlled cross between MR24 (with the *BsFLC1* deletion) and MAH (with a functional *BsFLC1* and requiring vernalization). From a self-fertilized hybrid F₁ we obtained 1,000 F₂ individuals. These seeds, along with 39 replicated individuals from each parental accession, were stratified in 4°C for 4 weeks, and seedlings were randomized and grown in the greenhouse without vernalization.

The presence or absence of the *BsFLC1* PCR amplicon is a dominant marker in this cross (both MAH homozygotes and heterozygotes have visible *BsFLC1* amplicon on agarose gel). We identified one polymorphic microsatellite co-dominant marker (marker JGI13175-36 at around 513 kb in Scaffold13175, 73 kb away from *BsFLC1*) and genotyped this marker in 384 F₂ individuals for the effect of *BsFLC1*.

Comparing the Expression Patterns and Molecular Evolution of *BsFLC* Copies

To characterize the possibly differentiated functions between these two *BsFLC* copies, we first investigated the expression profiles of both *BsFLC* copies as well as *FT* and *SOC1*, two floral pathway integrator genes that are directly inhibited by *FLC* in *Arabidopsis* (Helliwell et al., 2006). Multiple individuals from two accessions (MAH with functional *BsFLC1* and MR24 with *BsFLC1* deletion) were used in the experiment. Seeds were stratified in 4°C for 4 weeks. Plants were randomized and grown in the greenhouse for 7 weeks and separated into two experimental groups, either with or without 6-week vernalization in 4°C under 10-h days. One of the target genes, *FT*, exhibits circadian rhythm in gene expression in *Arabidopsis*, and under 16-h days its maximum expression is in leaves in the end of daytime (Yanovsky and Kay, 2002; Cheng and Wang, 2005; Kim et al., 2008). In addition, *FLC* and *SOC1* also express in *Arabidopsis* leaves (Lee and Lee, 2010). We therefore used leaf tissues collected around 22:00 h (the sunset of 16-h days in the greenhouse) for expression profiling. Young leaves from four individuals of each accession were collected from six time points (Supplementary Figure S1: BV – before vernalization, when plants were 7 weeks old; NV1 – no vernalization group 1, when plants were 9 weeks old, and MR24 flowered but MAH did not; NV2 – no vernalization group 2, when plants were 18 weeks old, and MR24 had mature and dehiscent siliques but MAH had first flowers; AV1 – after vernalization group 1, 1 week after vernalization when MR24 flowered but MAH did not; AV2 – after vernalization group 2, 3.5 weeks after vernalization when MR24 had flowers and fruits but MAH had only flowers; AV3 – after vernalization group 3, 6 weeks after vernalization when both accessions had only fruits). We used Sigma Spectrum™ Plant Total RNA Kit for RNA extraction and Thermo Scientific DyNamo cDNA Synthesis Kit for cDNA synthesis, resulting in 47 separate samples (six time points with two accessions and four individual plants each, where one sample was lost during RNA extraction).

qPCR from cDNA was performed for five genes (*BsFLC1*, *BsFLC2*, *FT*, *SOC1*, *ACT2*), with *ACT2* as the reference gene.

For each sample, gene expression was estimated with the ΔCt method:

$$\text{Relative expression to } ACT2 = 2^{-(\text{Ct}_{ACT2} - \text{Ct}_{\text{Gene}})},$$

where Ct_{ACT2} is the Ct value of *ACT2*, and Ct_{Gene} is the Ct value of target genes (*BsFLC1*, *BsFLC2*, *FT*, or *SOC1*). Using expression relative to *ACT2* as the response variable, statistical analyses (ANOVA) were conducted separately for each target gene, where genotype (MR24 or MAH), time point (BV, NV1, NV2, AV1, AV2, or AV3), and their interaction were treated as fixed effects in JMP 8 (SAS, Cary, NC, United States). To analyze the correlation of gene expression across the six time points, for each genotype separately, we further calculated the pairwise Spearman's rank correlation coefficient ρ between the four target genes.

To characterize the evolutionary history of the two *BsFLC* copies, we cloned and sequenced their full-length coding sequences in *B. stricta* accessions MR24, MAH, and LTM (accession for the reference genome). The sequences were deposited in GenBank under the accession numbers MH166767–MH166771. RNA was extracted with “Sigma Spectrum™ Plant Total RNA Kit” from rosette leaves, and cDNA was synthesized with “Life Technologies ThermoScript™ RT-PCR System for First-Strand cDNA Synthesis.” cDNA was amplified with conserved primers for both *FLC* copies (*FLC-H* in 5' UTR and *FLC-I* in 3' UTR, Supplementary Table S1 and Supplementary Figure S2) and cloned in “Thermo Scientific CloneJET PCR Cloning Kit.” Full-length coding sequences of other Brassicaceae species were obtained from GenBank: *Arabidopsis arenosa FLC1* (DQ167446) *FLC2* (DQ167444), *Arabidopsis halleri* (AB465585), *Arabidopsis suecica* (DQ167447), *Brassica napus FLC1* (AY036888) *FLC2* (AY036889) *FLC3* (AY036890) *FLC4* (AY036891) *FLC5* (AY036892), *Brassica rapa* (EF460819), *Brassica oleracea* (AY273161), *Capsella rubella* (JQ993010), *Cardamine flexuosa* (KC618318), *Eutrema wasabi* (HM639741), *Sinapis alba* (EF542803), *Thellungiella halophila* (AY957537), and *Raphanus sativus* (JX050205).

FLC coding sequences of Brassicaceae were examined in MEGA 4 (Tamura et al., 2007). A phylogenetic tree was reconstructed in MrBayes v3.1.2 (Ronquist and Huelsenbeck, 2003) with generalized time reversible model allowing invariable sites and gamma distribution for evolutionary rate difference among sites. Two independent runs were performed with eight chains. Trees were sampled every 1,000 generations, and the first 500 trees were discarded, leaving 1,500 trees for the tree topology summary. We only kept the tree topology and branch support from MrBayes (Rannala et al., 2012), and branch lengths were estimated with maximum likelihood in PAUP* 4.0b10 (Swofford, 2003) with the same model described above.

Our results showed that the two *BsFLC* copies in *B. stricta* were generated by a gene duplication event after *Boecheira* diverged from *Arabidopsis* and *Capsella*. We therefore used *Arabidopsis FLC* as the outgroup to analyze patterns of molecular evolution after *BsFLC* duplication. The ancestral sequence before gene duplication was reconstructed using PAML 4 (Yang, 2007), and synonymous and nonsynonymous substitutions were estimated with DnaSP v5 (Librado and Rozas, 2009). We

further used PAML 4 to test three hypotheses: (1) Whether the d_N/d_S values are significantly different between *BsFLC1* and *BsFLC2* (using the branch model, where the d_N/d_S values for the two genes are the same in the null hypothesis but different in the alternative hypothesis); (2) Whether the d_N/d_S value for *BsFLC2* is significantly higher than one, showing signs of positive selection (using the branch model, where the d_N/d_S value for *BsFLC2* was restricted to one in the null hypothesis but unrestricted in the alternative hypothesis); and (3) Was there any codon site under strong positive selection on the *BsFLC2* branch (the branch-site model, Zhang et al., 2005).

We used custom R scripts (R Core Team, 2014) to investigate whether amino acid substitutions accumulated on *BsFLC2* after gene duplication occurred randomly on the protein or clustered in a specific functional domain. Among the total of 196 amino acids in this protein, we randomly draw nine “substitutions” (*BsFLC2* accumulated nine substitutions after duplication) and recorded the number of substitutions inside the K-box domain. This procedure was repeated 1,000 times, and the observed value was compared with the distribution of 1,000 K-box-specific substitutions under the assumption that substitutions occurred randomly along the protein.

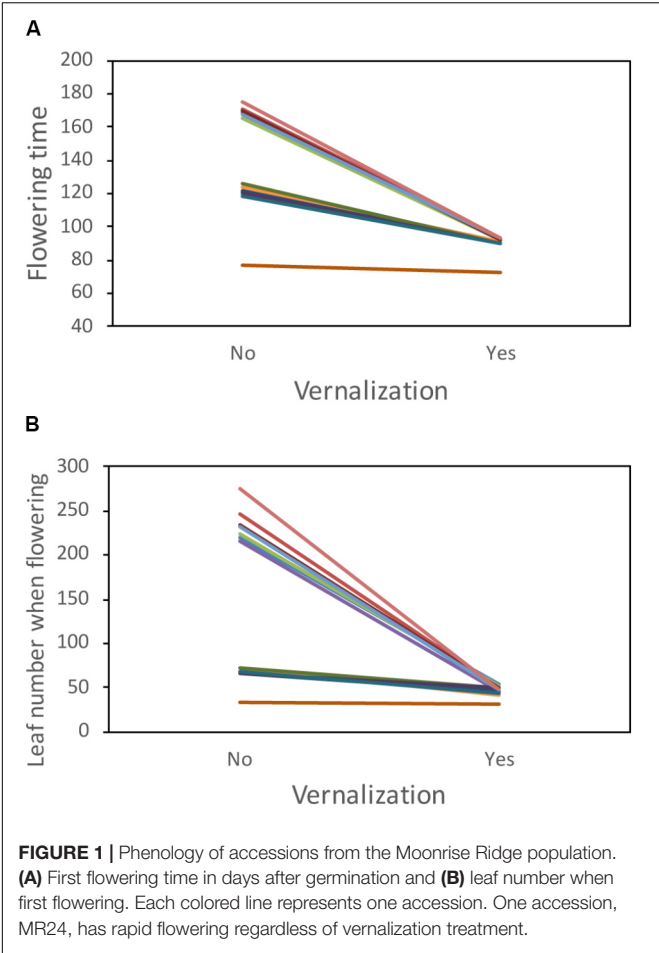
RESULTS

High Genetic Variation for Phenology Within a Population

Accessions from the Moonrise Ridge population showed highly significant differences in their phenology and response to vernalization treatment. Specifically, treatment and accession-by-treatment interaction had highly significant effects on both flowering time and leaf number at first flower (Figure 1 and Table 1). Notably, while vernalization accelerated flowering in most accessions, the accession MR24 exhibited rapid flowering regardless of the vernalization treatment. This lack of vernalization requirement in MR24 is unusual in *B. stricta*, and we focused on the genetic mechanism underlying this trait.

Candidate Gene Profiling Identified *BsFLC1* as Possible Causal Gene

Since the flowering of MR24 was not affected by vernalization, we investigated homologs for the candidate genes *FLC* (the key gene underlying the vernalization requirement in *Arabidopsis*) in the rapid-flowering MR24 and slow-flowering MR7A accessions. Previous work identified two *FLC* copies in *Boecheira* (Schranz et al., 2007), and these two copies showed distinct patterns between MR24 and MR7A (Figure 2): while *BsFLC2* is present and expressed in both accessions, the *BsFLC1* locus was only amplified in the slow-flowering MR7A but not the rapid-flowering MR24 accession. To test whether this lack of *BsFLC1* amplification in MR24 was due to deletion of *BsFLC1* or mutation in primer binding sites, three forward and three reverse primers were used to amplify *BsFLC1*, and products from all nine primer combinations were successfully amplified in MR7A but not in



MR24. We therefore conclude that *BsFLC1* was lost in the rapid-flowering MR24 accession. *FLC* in *Arabidopsis* functions as a floral repressor and is suppressed by vernalization, and the lack of *BsFLC1* in MR24 is consistent with its rapid-flowering phenotype regardless of vernalization treatment.

We used the Nanopore long-read sequencing technology to confirm and investigate the extent of this deletion. We identified three reads supporting the chromosomal deletion. Further PCR and Sanger sequencing from primer pairs spanning the deletion breakpoint confirmed this finding (Supplementary Figure S3). This deletion covers *BsFLC1* and is about 19-kb long, ranging from 435.5 to 545.6 kb on Scaffold13175 of the reference genome. In addition, another upstream ~9 kb deletion was identified by a long read (Supplementary Figure S3). This region might be an accession-specific insertion in the LTM reference genome, as we did not find this indel in the alignment between this long Nanopore read and the *Boechera retrofracta* (a sister species of *B. stricta*) assembly (Kliver et al., 2018).

The Effect of *BsFLC1* on Flowering Time

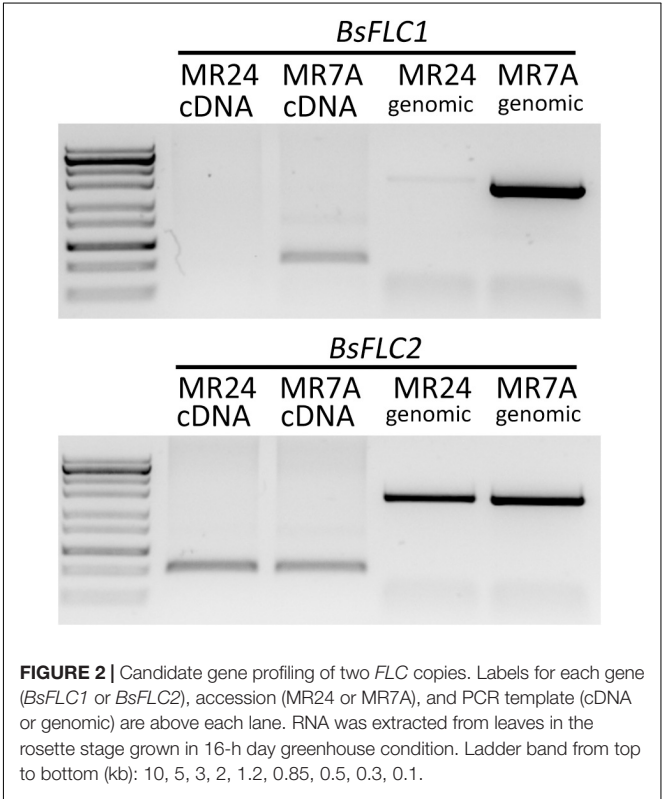
To estimate the effect of the *BsFLC1* deletion, flowering time was measured in 1,000 F₂ plants (Supplementary Figure S4) from a cross between parents MR24 (with *BsFLC1* deletion) and MAH (with functional *BsFLC1*), and 384 plants were genotyped

for a microsatellite marker (JGI13175-36) ~73 kb from *BsFLC1* in the reference genome and ~58 kb from the deletion in the MR24 genome. From the re-sequencing of 159 F₆ recombinant inbred lines (Lee et al., 2017), no recombination event was observed between *BsFLC1* and this microsatellite marker. The microsatellite had highly significant association with flowering time (Figure 3, $P = 10^{-61}$), accounting for about 70% of phenotypic variation in the F₂ plants, and the number of functional *BsFLC1* alleles showed roughly additive effects on flowering time (Figure 3), consistent with other Brassicaceae species (Schrantz et al., 2002). Our results are therefore consistent with the hypothesis that the *BsFLC1* deletion is associated with the rapid phenology in MR24 accession.

TABLE 1 | Analysis of variance of phenology in the Moonrise Ridge population.

Factor (effect)	Flowering time		Leaf number when flowering	
	Statistic ^a	P-value	Statistic	P-value
Vernalization (fixed)	$F_{1,13} = 54.71$	<0.001	$F_{1,13} = 18.47$	<0.001
Accession (random)	$\chi^2_1 = 1.01$	0.315	$\chi^2_1 = 0.05$	0.822
Accession	$\chi^2_1 = 980.27$	<0.001	$\chi^2_1 = 545.26$	<0.001
*Vernalization (random)				
Block (random)	$\chi^2_1 = 0$	1.000	$\chi^2_1 = 6.60$	0.010

^aSignificance of random effects was determined by likelihood ratio test between models with or without these effects. Subscript numbers after *F* denote the numerator and denominator degrees of freedom, and the subscript number after χ^2 denotes degrees of freedom for likelihood ratio test.



Expression Patterns of *BsFLC1*, *BsFLC2*, *FT*, and *SOC1*

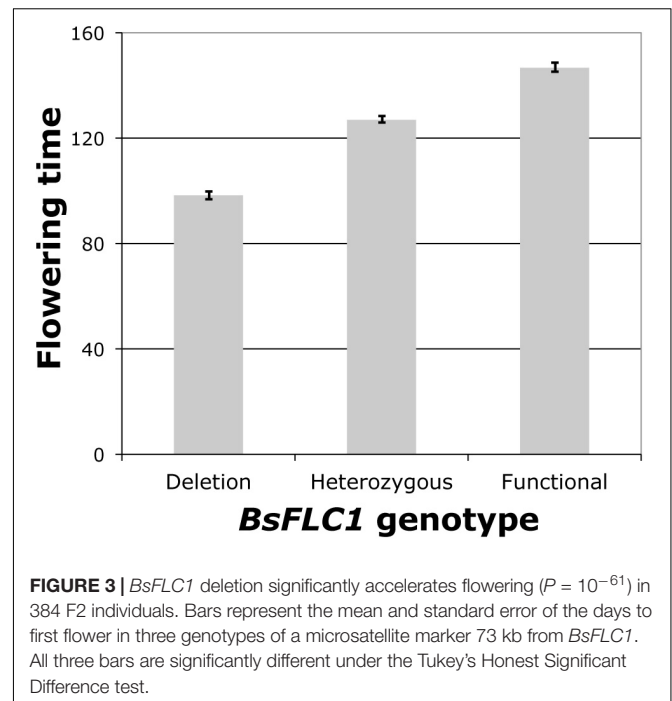
While the previous F₂ linkage analysis showed the genomic region containing *BsFLC1* deletion is strongly associated with rapid phenology, many genes exist in this genomic region. To find further support that *BsFLC1* is associated with flowering time but *BsFLC2* is not, we used qPCR to characterize the expression pattern of the two genes, as well as *FT* and *SOC1*, two floral pathway integrators that are directly inhibited by *FLC* in *Arabidopsis* (Helliwell et al., 2006). Replicated individuals from two parental accessions (MR24 and MAH) were sampled across six time points with or without vernalization (Supplementary Figure S1). Due to the deletion, *BsFLC1* in MR24 had no qPCR signal (Figure 4A). Both genes have similar levels of expression during the rosette stage before vernalization. The expression of *BsFLC1* in MAH associated strongly with plant stage: high expression in vegetative stage but low in reproductive stage (Figure 4A). In rosettes, *BsFLC1* expression was suppressed by vernalization (Figure 4A from time point BV to AV1) and decreased through time even without vernalization treatment (Figure 4A from time point BV to NV1 to NV2). On the other hand, the expression pattern of *BsFLC2* had no clear association with plant stages. In fact, the high expression of *BsFLC2* in some reproductive stages suggested this gene did not retain the ancestral function to suppress flowering (Figure 4B). *BsFLC2* was also suppressed by vernalization in either accessions with or without *BsFLC1*, although the magnitude of suppression is not as much as *BsFLC1* (Figure 4B from time point BV to AV1).

In the MAH accession with functional *BsFLC1*, the expression pattern of two downstream floral integrators, *FT* and *SOC1*, also associated strongly with plant stages (Figures 4C,D). Opposite to *BsFLC1*, both genes had higher expression during reproductive but lower expression in vegetative stages, although the pattern in *FT* was less clear due to its generally lower expression. In addition, the correlation between *BsFLC1* and *SOC1* expression was strongly negative across these six time points (Spearman's $\rho = -0.94$, $P = 0.005$, Table 2). In the MR24 accession with *BsFLC1* deletion, however, the association between *FT* or *SOC1* and plant stage was disrupted (Figures 4C,D). *BsFLC2* had no significant correlation with the two downstream genes in either MAH or MR24 (Table 2).

In summary, *BsFLC1* in *B. stricta* may have retained the ancestral flowering-related function: its expression has high association with the vegetative/reproductive plant stages, is reduced after vernalization treatments, and is correlated with downstream flowering genes. In contrast, *BsFLC2* may have diverged in function, although its expression was still suppressed by vernalization.

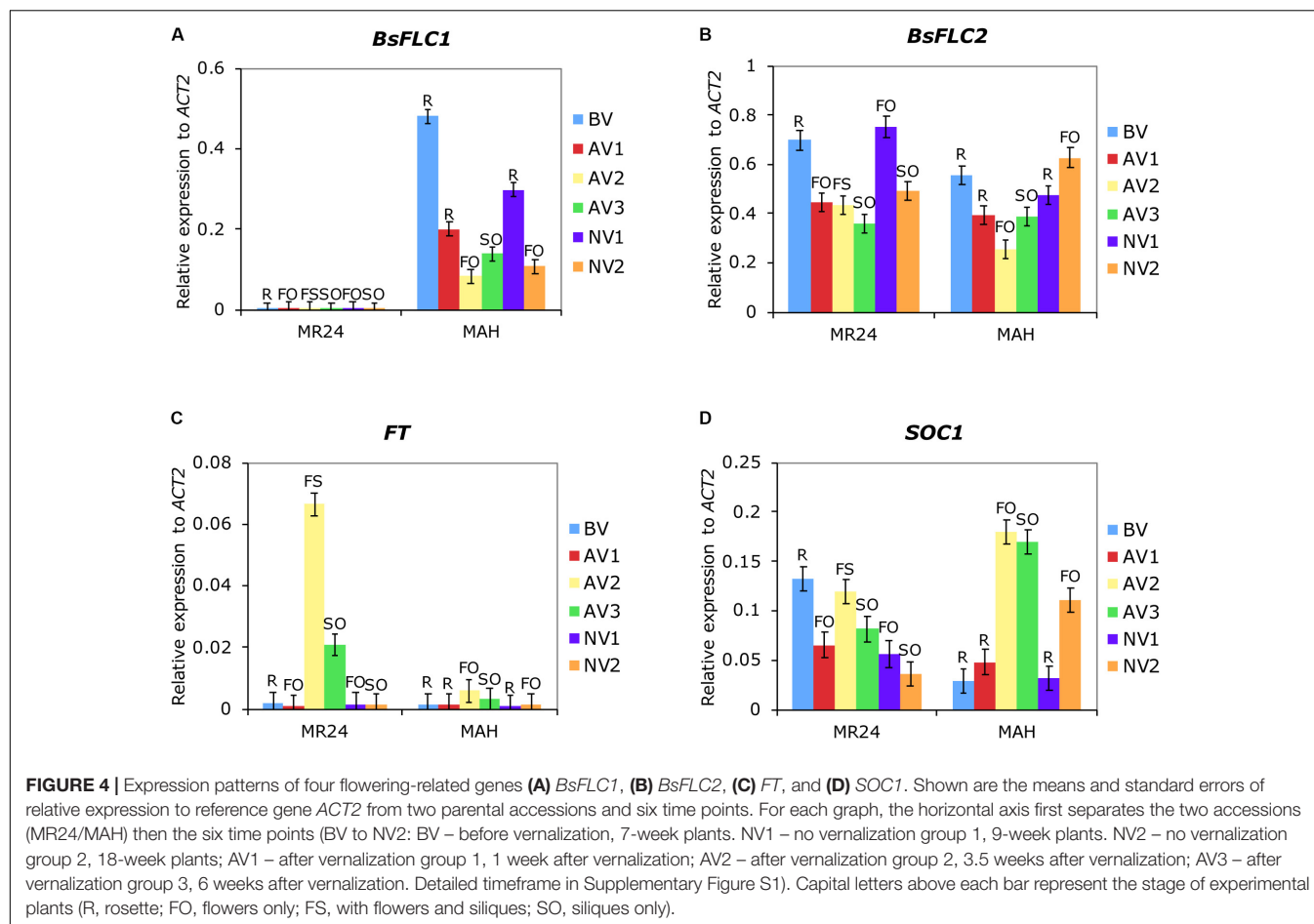
Evolution of the Two FLC Copies After Gene Duplication

Phylogenetic reconstruction of *FLC* homologs across Brassicaceae showed clearly that the gene duplication event happened after the *Boechera-Arabidopsis* divergence (Figure 5). We therefore focused our analysis on *BsFLC1* and *BsFLC2* within *Boechera*, using *Arabidopsis FLC* as the outgroup.



After duplication, *BsFLC1* accumulated three substitutions (one synonymous and two non-synonymous) while *BsFLC2* accumulated 10 substitutions (one synonymous, eight non-synonymous, and a three-codon deletion). Using all substitution types, Tajima's relative rate test (Tajima, 1993) identified no strong evolutionary rate difference between the two copies ($\chi^2_1 = 3.769$, $P = 0.052$). Considering only non-synonymous substitutions, the evolutionary rate of *BsFLC2* is significantly higher than *BsFLC1* ($\chi^2_1 = 4.455$, $P = 0.036$). In addition, *BsFLC2* accumulated more non-synonymous substitutions than *BsFLC1*, resulting in much higher $d_N:d_S$ ratio (0.608 for *BsFLC1* and 2.460 for *BsFLC2*, without considering the three-codon deletion in *BsFLC2*, Table 3). However, statistical tests using PAML showed that: (1) The $d_N:d_S$ ratios do not differ significantly between *BsFLC1* and *BsFLC2* ($P = 0.421$); (2) The $d_N:d_S$ ratio for *BsFLC2* is not significantly different from 1.0 ($P = 0.235$); and (3) There is little evidence that specific codons on the *BsFLC2* lineage were under positive selection ($P = 0.236$).

The distribution of amino acid substitutions after gene duplication is not random (Supplementary Figure S5). The two amino acid substitutions on *BsFLC1* do not have an obvious spatial concentration. In contrast, all *BsFLC2* changes are concentrated on or near the K-box domain. Assuming amino acid substitutions happened independently and randomly on *BsFLC2*, it is highly unlikely that eight out of nine amino acid changes would be located inside the K-box domain ($P = 0.006$), as only six out of one thousand resampling trials gave eight or more changes within the K-box domain. The types of amino acid substitutions also differed between the two *BsFLC* copies. While both *BsFLC1* substitutions and most substitutions on *BsFLC2* resulted in the change of amino acids with physicochemical distances less than 56 (Grantham, 1974), the *BsFLC2* lineage contained



three nearby amino acid substitutions that greatly altered the physicochemical properties on these positions (distance of 99, 109, and 98 for amino acid positions 116, 118, and 121, respectively, Supplementary Figure S5).

DISCUSSION

Repeated Evolution of Flowering Time Variation Through Change of FLC

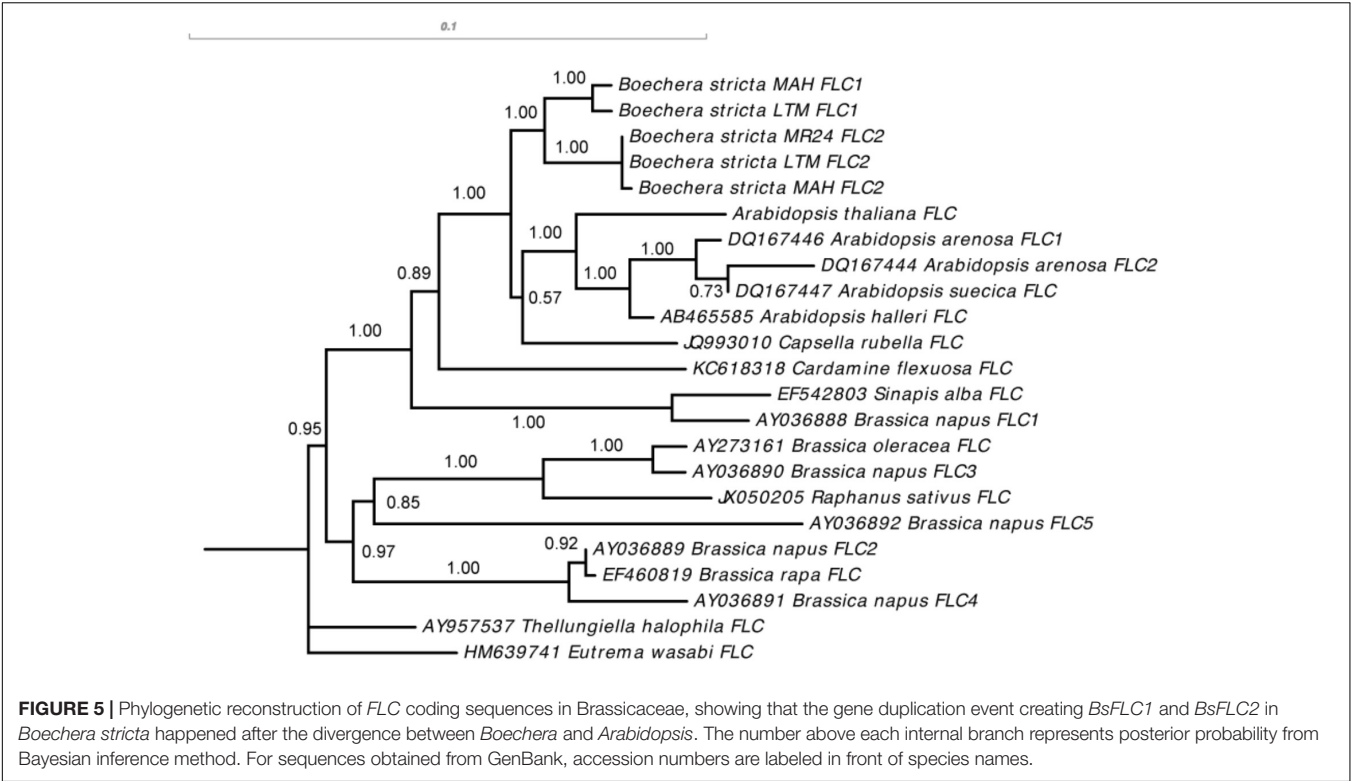
While phenotypic changes can be achieved by segregation and recombination of standing genetic variants, mutations play an important role in generating novel phenotypes. Novel mutations differ in size (from Mendelian to polygenes) and direction (from advantageous, neutral, to deleterious) of their effects (Orr, 2005; Eyre-Walker and Keightley, 2007). Due to their unique roles in development or positions in biochemical pathways, genes with large effect may often be the targets of repeated evolution for similar phenotypes (Martin and Orgogozo, 2013). The *FLC* gene constitutes one example. In addition to *BsFLC1* in *B. stricta*, independent mutations in *FLC* homologs cause heritable variation in vernalization requirement of flowering time in other Brassicaceae species (Alonso-Blanco and Méndez-Vigo, 2014) such as *A. thaliana*

(Michaels et al., 2003; Lempe et al., 2005; Shindo et al., 2005; Werner et al., 2005; Méndez-Vigo et al., 2011), *Arabidopsis lyrata* (Kemi et al., 2013), *Arabis alpina* (Wang et al., 2009; Albani et al., 2012), *Capsella rubella* (Guo et al., 2012; Yang et al., 2018), *Brassica napus* (Tadege et al., 2001), *Brassica oleracea* (Okazaki et al., 2007), and *Brassica rapa* (Schranz et al., 2002). Recently, *ODDSOC2*, the *FLC/MAF* ortholog in monocots, was also shown to be associated with the vernalization requirement of flowering in *Brachypodium distachyon* (Sharma et al., 2016).

Here we studied the genetic architecture of rapid phenology in *B. stricta* and investigated the two *FLC* homologs in *B. stricta*. We showed that the gene duplication event is specific to *Boechea* (Figure 5), and previous data have

TABLE 2 | Pairwise correlation (Spearman's ρ , with *P*-values in parentheses) between the expression patterns of four genes from six time points in two accessions (upper triangle, MAH; lower triangle, MR24, the accession without *BsFLC1*).

<i>BsFLC1</i>	0.43 (0.397)	−0.43 (0.397)	−0.94 (0.005)
–	<i>BsFLC2</i>	−0.71 (0.111)	−0.66 (0.156)
–	−0.43 (0.397)	<i>FT</i>	0.60 (0.208)
–	−0.26 (0.623)	0.66 (0.156)	<i>SOC1</i>



identified both *FLC* copies in multiple *Boechera* species (Schrantz, unpublished data). *BsFLC1* is the homolog of *A. thaliana FLC* in the syntenic region. This genomic region controls large variation in flowering time (Figure 3), and accessions with the *BsFLC1* deletion (Figure 2 and Supplementary Figure S3) flowered rapidly regardless of vernalization treatment (Figure 1). We recognize that our vernalization treatment changed both temperature and photoperiod within biologically realistic limits, and there may be other genetic mechanisms affecting such response. In addition, the expression pattern of *BsFLC1* is similar to other Brassicaceae plants and highly associated with plant vegetative/reproductive stages and the expression pattern of downstream flowering genes (Figure 4 and Table 2). In the annual plant *A. thaliana*, *FLC* expression is stably repressed epigenetically after vernalization and “reset” only during embryogenesis (Sheldon et al., 2008; Choi et al., 2009; Berry and Dean, 2015). In perennial plants such as *Arabidopsis alpina*, expression of *FLC* is also repressed during but restored after vernalization, allowing plants to return to the vegetative stage (Wang et al., 2009; Kiefer et al., 2017). The expression

pattern of *BsFLC1* is similar to the latter, consistent with the perennality of *B. stricta*. Our analyses suggest that after gene duplication, *BsFLC1* likely retained the ancestral function to suppress flowering, and the evolution of rapid phenology in *B. stricta* accession MR24 happened through the deletion of *BsFLC1*.

Change of *BsFLC2* Function Enables Evolution of Rapid Phenology Through *BsFLC1* Deletion

The other gene, *BsFLC2*, is the unlinked and duplicated paralog. The mismatch between *BsFLC2*’s expression pattern and plant vegetative/reproductive stage, as well as its strong expression in the fast-flowering MR24 accession, suggest it does not retain the ancestral function to control flowering (Figure 4). As a result, one deletion event in *BsFLC1* is enough to generate the loss-of-vernalization-requirement phenotype in MR24.

As to how or when *BsFLC2* lost the flowering-related functions, there are several possibilities (Introduction). First, only

TABLE 3 | Synonymous and non-synonymous substitutions in *BsFLC1* and *BsFLC2* after gene duplication.

Gene	S differences	S sites	d_S	N differences	N sites	d_N	d_N/d_S
<i>BsFLC1</i>	1	134.33	0.0074	2	444.67	0.0045	0.6081
<i>BsFLC2</i>	1	134.50	0.0074	8	444.50	0.0182	2.4595

This table compares the difference between each gene and the ancestral sequence before gene duplication. Columns: S differences, number of synonymous differences; S sites, number of synonymous sites; d_S , S differences/S sites; N differences, number of nonsynonymous differences; N sites, number of nonsynonymous sites; d_N , N differences/N sites.

part of the gene was duplicated, and *BsFLC2* is a truncated non-functional gene. Our results show that *BsFLC2* is likely to be functional given its intact and in-frame coding sequence, the conserved MADS-box domain (Supplementary Figure S5), the retention of several kbs of ancestral sequence upstream and downstream (Supplementary Figure S6), and the ability to be suppressed by vernalization.

Second, the duplication event may disrupt the upstream or downstream regulatory sequence, causing *BsFLC2* to be expressed in completely different tissues or times, thereby altering the biological function of *BsFLC1*. We observe that the expression patterns between the two copies are qualitatively similar (Figure 4): both copies had similar levels of expression in young rosette leaves (where and when the flowering-suppression function happened) and were suppressed by vernalization. The duplication event therefore does not seem to dramatically change the expression pattern of *BsFLC2*. We note, however, that there are subtle differences between the two copies: *BsFLC2* expression was not suppressed by vernalization as much as *BsFLC1*, and it was highly expressed during the flowering stage for plants without the vernalization treatment (Figure 4). Since the precise regulation of *A. thaliana FLC* requires non-coding regions within and flanking the gene (Michaels and Amasino, 1999; Swiezewski et al., 2009; Sun et al., 2013; Csorba et al., 2014), *BsFLC2* might not retain enough ancestral regulatory sequences (Supplementary Figure S6). Therefore, it is possible that the duplication event may cause *BsFLC2* to exhibit slightly different expression patterns than *BsFLC1*, although it remains unclear whether such difference was directly caused by or happened after the gene duplication event.

While the two previous possibilities concern the duplication event directly disrupting the coding sequence or regulatory region of *BsFLC2*, our results are also consistent with the third hypothesis: the coding and regulatory functions of *BsFLC2* remained largely unaffected by the gene duplication event, and *BsFLC2* diverged in protein function afterward. We observed protein sequence evolution in *BsFLC2* that cannot be completely explained by neutral evolution following gene duplication. Depending on the methods used, we sometimes obtained mixed results regarding whether the evolutionary rate differs significantly between *BsFLC1* and *BsFLC2*. It should be noted that these results might be in part due to the lack of statistical power, given the limited number of nucleotide substitutions after this very recent gene duplication event. In addition, these d_N/d_S -based methods ignored the three-codon deletion in *BsFLC2*, which removed three amino acids in a functionally important domain (below).

The *FLC* protein is a MIKC-type protein, consisting of the MADS (M-), intervening (I-), keratin-like (K-), and C-terminal (C-) domains (Theissen et al., 1996; Kaufmann et al., 2005). The major function of MADS-box is DNA-binding (Kaufmann et al., 2005), and in *FLC* this domain is associated with chromatin interactions of *FT* and *SOC1* (Helliwell et al., 2006), two downstream floral pathway integrator genes. The other domains, especially the K-box domain, are associated with dimerization or protein-protein interaction (Kaufmann et al., 2005) and may be important to form the functional protein complex, as multiple *FLC* proteins were

identified in the same multimeric protein complex (Helliwell et al., 2006). All amino acid substitutions in *BsFLC2* are concentrated near the K-box (Supplementary Figure S5), suggesting that the *BsFLC2* protein may have undergone functional change after gene duplication, interacting with other proteins.

In this study, we investigate the interplay between loss-of-function evolution and gene duplication. We show that *BsFLC1* and *BsFLC2* differ slightly in expression patterns, which might be caused either directly by the gene duplication event or by later mutations. More importantly, after gene duplication *BsFLC2* accumulated amino acid substitutions in a speed higher than neutral evolution and in a region more concentrated than random expectation, suggesting directional selection driving its protein sequence evolution. We propose such changes alter protein functions, allowing a single deletion event encompassing *BsFLC1* to create a novel loss-of-vernalization-requirement phenotype in the MR24 accession despite the continual expression of *BsFLC2* in rosette leaves.

DATA ACCESSIBILITY

DNA sequences: GenBank MH166767-MH166771. DNA sequence alignment, phenotypic values, and qPCR results: uploaded as Supplementary Materials.

AUTHOR CONTRIBUTIONS

C-RL, MS, and TM-O designed the study. C-RL and J-WH conducted experiments. C-RL analyzed data and wrote the manuscript with help from all authors.

FUNDING

This work is supported by the Ministry of Science and Technology of Taiwan (105-2311-B-002-040-MY2 and 107-2636-B-002-004 to C-RL), the US National Institutes of Health (R01 GM086496 to TM-O), the US National Science Foundation (EF-0723447 to TM-O), and Netherlands Science Foundation (NWO) VIDI and Ecogenomics Grants to MS.

ACKNOWLEDGMENTS

We are grateful to Computer and Information Networking Center, National Taiwan University for the support of high-performance computing facilities.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.01078/full#supplementary-material>

REFERENCES

- Albani, M. C., Castaings, L., Wötzel, S., Mateos, J. L., Wunder, J., Wang, R., et al. (2012). PEP1 of *Arabidopsis alpinia* is encoded by two overlapping genes that contribute to natural genetic variation in perennial flowering. *PLoS Genet.* 8:e1003130. doi: 10.1371/journal.pgen.1003130
- Alonso-Blanco, C., and Méndez-Vigo, B. (2014). Genetic architecture of naturally occurring quantitative traits in plants: an updated synthesis. *Curr. Opin. Plant Biol.* 18, 37–43. doi: 10.1016/j.pbi.2014.01.002
- Anderson, J. T., Lee, C.-R., and Mitchell-Olds, T. (2011). Life history QTLs and natural selection on flowering time in *Boechera stricta*, a perennial relative of *Arabidopsis*. *Evolution* 65, 771–787. doi: 10.1111/j.1558-5646.2010.01175.x
- Anderson, J. T., Lee, C.-R., and Mitchell-Olds, T. (2014). Strong selection genome-wide enhances fitness trade-offs across environments and episodes of selection. *Evolution* 68, 16–31. doi: 10.1111/evo.12259
- Anderson, J. T., Lee, C.-R., Rushworth, C., Colautti, R. I., and Mitchell-Olds, T. (2012). Genetic tradeoffs and conditional neutrality contribute to local adaptation. *Mol. Ecol.* 22, 699–708. doi: 10.1111/j.1365-294X.2012.05522.x
- Berry, S., and Dean, C. (2015). Environmental perception and epigenetic memory: mechanistic insight through FLC. *Plant J.* 83, 133–148. doi: 10.1111/tpj.12869
- Cheng, X. F., and Wang, Z. Y. (2005). Overexpression of COL9, a CONSTANS-LIKE gene, delays flowering by reducing expression of CO and FT in *Arabidopsis thaliana*. *Plant J.* 43, 758–768. doi: 10.1111/j.1365-313X.2005.02491.x
- Choi, J., Hyun, Y., Kang, M. J., In Yun, H., Yun, J. Y., Lister, C., et al. (2009). Resetting and regulation of FLOWERING LOCUS C expression during *Arabidopsis* reproductive development. *Plant J.* 57, 918–931. doi: 10.1111/j.1365-313X.2008.03776.x
- Csorba, T., Questa, J. I., Sun, Q., and Dean, C. (2014). Antisense COOLAIR mediates the coordinated switching of chromatin states at FLC during vernalization. *Proc. Natl. Acad. Sci. U.S.A.* 111, 16160–16165. doi: 10.1073/pnas.1419030111
- Doyle, J. J., and Doyle, J. L. (1987). A rapid isolation procedure for small DNA quantities of fresh leaf tissues. *Phytochem. Bull.* 19, 11–15.
- Eyre-Walker, A., and Keightley, P. D. (2007). The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* 8:610. doi: 10.1038/nrg2146
- Gazzani, S., Gendall, A. R., Lister, C., and Dean, C. (2003). Analysis of the molecular basis of flowering time variation in *Arabidopsis* accessions. *Plant Physiol.* 132, 1107–1114. doi: 10.1104/pp.103.021212
- Grantham, R. (1974). Amino acid difference formula to help explain protein evolution. *Science* 185, 862–864. doi: 10.1126/science.185.4154.862
- Guo, Y.-L., Todesco, M., Hagmann, J., Das, S., and Weigel, D. (2012). Independent FLC mutations as causes of flowering-time variation in *Arabidopsis thaliana* and *Capsella rubella*. *Genetics* 192, 729–739. doi: 10.1534/genetics.112.143958
- Helliwell, C. A., Wood, C. C., Robertson, M., James Peacock, W., and Dennis, E. S. (2006). The *Arabidopsis* FLC protein interacts directly in vivo with SOC1 and FT chromatin and is part of a high-molecular-weight protein complex. *Plant J.* 46, 183–192. doi: 10.1111/j.1365-313X.2006.02686.x
- Johanson, U., West, J., Lister, C., Michaels, S., Amasino, R., and Dean, C. (2000). Molecular analysis of FRIGIDA, a major determinant of natural variation in *Arabidopsis* flowering time. *Science* 290, 344–347. doi: 10.1126/science.290.5490.344
- Kaufmann, K., Melzer, R., and Theißen, G. (2005). MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. *Gene* 347, 183–198. doi: 10.1016/j.gene.2004.12.014
- Kemi, U., Niittyvuopio, A., Toivainen, T., Pasanen, A., Quilót-Turion, B., Holm, K., et al. (2013). Role of vernalization and of duplicated FLOWERING LOCUS C in the perennial *Arabidopsis lyrata*. *New Phytol.* 197, 323–335. doi: 10.1111/j.1469-8137.2012.04378.x
- Kiefer, C., Severing, E., Karl, R., Bergonzi, S., Koch, M., Tresch, A., et al. (2017). Divergence of annual and perennial species in the Brassicaceae and the contribution of cis-acting variation at FLC orthologues. *Mol. Ecol.* 26, 3437–3457. doi: 10.1111/mec.14084
- Kim, S. Y., Yu, X., and Michaels, S. D. (2008). Regulation of CONSTANS and FLOWERING LOCUS T expression in response to changing light quality. *Plant Physiol.* 148, 269–279. doi: 10.1104/pp.108.122606
- Kliver, S., Rayko, M., Komissarov, A., Bakin, E., Zhernakova, D., Prasad, K., et al. (2018). Assembly of the *Boechera retrofracta* genome and evolutionary analysis of apomixis-associated genes. *Genes* 9:185. doi: 10.3390/genes9040185
- Le Corre, V., Roux, F., and Reboud, X. (2002). DNA polymorphism at the FRIGIDA gene in *Arabidopsis thaliana*: extensive nonsynonymous variation is consistent with local selection for flowering time. *Mol. Biol. Evol.* 19, 1261–1271. doi: 10.1093/oxfordjournals.molbev.a004187
- Lee, C. R., Anderson, J. T., and Mitchell-Olds, T. (2014). Unifying genetic canalization, genetic constraint, and genotype-by-environment interaction: QTL by genomic background by environment interaction of flowering time in *Boechera stricta*. *PLoS Genet.* 10:e1004727. doi: 10.1371/journal.pgen.1004727
- Lee, C.-R., and Mitchell-Olds, T. (2011). Quantifying effects of environmental and geographical factors on patterns of genetic differentiation. *Mol. Ecol.* 20, 4631–4642. doi: 10.1111/j.1365-294X.2011.05310.x
- Lee, C.-R., and Mitchell-Olds, T. (2013). Complex trait divergence contributes to environmental niche differentiation in ecological speciation of *Boechera stricta*. *Mol. Ecol.* 22, 2204–2217. doi: 10.1111/mec.12250
- Lee, C.-R., Wang, B., Mojica, J. P., Mandáková, T., Prasad, K. V., Goicoechea, J. L., et al. (2017). Young inversion with multiple linked QTLs under selection in a hybrid zone. *Nat. Ecol. Evol.* 1:0119.
- Lee, J., and Lee, I. (2010). Regulation and function of SOC1, a flowering pathway integrator. *J. Exp. Bot.* 61, 2247–2254. doi: 10.1093/jxb/erq098
- Lempe, J., Balasubramanian, S., Sureshkumar, S., Singh, A., Schmid, M., Weigel, D., et al. (2015). Diversity of flowering responses in wild *Arabidopsis thaliana* strains. *PLoS Genet.* 11:e1004666. doi: 10.1371/journal.pgen.1004666
- Li, P., Filiault, D., Box, M. S., Kerdaffrec, E., van Oosterhout, C., Wilczek, A. M., et al. (2014). Multiple FLC haplotypes defined by independent cis-regulatory variation underpin life history diversity in *Arabidopsis thaliana*. *Genes Dev.* 28, 1635–1640. doi: 10.1101/gad.245993.114
- Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452. doi: 10.1093/bioinformatics/btp187
- Martin, A., and Orgogozo, V. (2013). The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evolution* 67, 1235–1250. doi: 10.1111/evo.12081
- Méndez-Vigo, B., Picó, F. X., Ramiro, M., Martínez-Zapater, J. M., and Alonso-Blanco, C. (2011). Altitudinal and climatic adaptation is mediated by flowering traits and FRI, FLC, and PHYC genes in *Arabidopsis*. *Plant Physiol.* 157, 1942–1955. doi: 10.1104/pp.111.183426
- Michaels, S. D., and Amasino, R. M. (1999). FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* 11, 949–956. doi: 10.1105/tpc.11.5.949
- Michaels, S. D., He, Y., Scortecci, K. C., and Amasino, R. M. (2003). Attenuation of FLOWERING LOCUS C activity as a mechanism for the evolution of summer-annual flowering behavior in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 100, 10102–10107. doi: 10.1073/pnas.1531467100
- Mungruía-Rosas, M. A., Ollerton, J., Parra-Tabla, V., and De-Nova, J. A. (2011). Meta-analysis of phenotypic selection on flowering phenology suggests that early flowering plants are favoured. *Ecol. Lett.* 14, 511–521. doi: 10.1111/j.1461-0248.2011.01601.x
- Okazaki, K., Sakamoto, K., Kikuchi, R., Saito, A., Togashi, E., Kuginuki, Y., et al. (2007). Mapping and characterization of FLC homologs and QTL analysis of flowering time in *Brassica oleracea*. *Theor. Appl. Genet.* 114:595. doi: 10.1007/s00122-006-0460-6
- Orr, H. A. (2005). The genetic theory of adaptation: a brief history. *Nat. Rev. Genet.* 6, 119–127. doi: 10.1038/nrg1523
- R Core Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available at: <http://www.R-project.org/>
- Rannala, B., Zhu, T., and Yang, Z. (2012). Tail paradox, partial identifiability, and influential priors in Bayesian branch length inference. *Mol. Biol. Evol.* 29, 325–335. doi: 10.1093/molbev/msr210
- Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., et al. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26. doi: 10.1038/nbt.1754
- Ronquist, F., and Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574. doi: 10.1093/bioinformatics/btg180

- Schranz, M. E., Quijada, P., Sung, S.-B., Lukens, L., Amasino, R., and Osborn, T. C. (2002). Characterization and effects of the replicated flowering time gene FLC in *Brassica rapa*. *Genetics* 162, 1457–1468.
- Schranz, M. E., Windsor, A. J., Song, B.-H., Lawton-Rauh, A., and Mitchell-Olds, T. (2007). Comparative genetic mapping in *Boechera stricta*, a close relative of *Arabidopsis*. *Plant Physiol.* 144, 286–298. doi: 10.1104/pp.107.096685
- Sharma, N., Ruelens, P., Dhauw, M., Maggen, T., Dochy, N., Torfs, S., et al. (2016). A Flowering locus C homolog is a vernalization-regulated repressor in *Brachypodium* and is cold-regulated in wheat. *Plant Physiol.* 173, 1301–1315. doi: 10.1104/pp.16.01161
- Sheldon, C. C., Finnegan, E. J., Rouse, D. T., Tadege, M., Bagnall, D. J., Helliwell, C. A., et al. (2000). The control of flowering by vernalization. *Curr. Opin. Plant Biol.* 3, 418–422. doi: 10.1016/S1369-5266(00)00106-0
- Sheldon, C. C., Hills, M. J., Lister, C., Dean, C., Dennis, E. S., and Peacock, W. J. (2008). Resetting of FLOWERING LOCUS C expression after epigenetic repression by vernalization. *Proc. Natl. Acad. Sci. U.S.A.* 12, 2214–2219. doi: 10.1073/pnas.0711453105
- Shindo, C., Aranzana, M. J., Lister, C., Baxter, C., Nicholls, C., Nordborg, M., et al. (2005). Role of FRIGIDA and FLOWERING LOCUS C in determining variation in flowering time of *Arabidopsis*. *Plant Physiol.* 138, 1163–1173. doi: 10.1104/pp.105.061309
- Sović, I., Šikić, M., Wilm, A., Fenlon, S. N., Chen, S., and Nagarajan, N. (2016). Fast and sensitive mapping of nanopore sequencing reads with GraphMap. *Nat. Commun.* 7:11307. doi: 10.1038/ncomms11307
- Sun, Q., Csorba, T., Skourti-Stathaki, K., Proudfoot, N. J., and Dean, C. (2013). R-loop stabilization represses antisense transcription at the *Arabidopsis* FLC locus. *Science* 340, 619–621. doi: 10.1126/science.1234848
- Swiezewski, S., Liu, F., Magusin, A., and Dean, C. (2009). Cold-induced silencing by long antisense transcripts of an *Arabidopsis* polycomb target. *Nature* 462:799. doi: 10.1038/nature08618
- Swofford, D. L. (2003). *PAUP*: Phylogenetic Analysis Using Parsimony (*and other methods)*. Sunderland, MA: Sinauer Associates.
- Tadege, M., Sheldon, C. C., Helliwell, C. A., Stoutjesdijk, P., Dennis, E. S., and Peacock, W. J. (2001). Control of flowering time by FLC orthologues in *Brassica napus*. *Plant J.* 28, 545–553. doi: 10.1046/j.1365-313X.2001.01182.x
- Tajima, F. (1993). Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* 135, 599–607.
- Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) Software Version 4.0. *Mol. Biol. Evol.* 24, 1596–1599. doi: 10.1093/molbev/msm092
- Theißen, G., Kim, J. T., and Saedler, H. (1996). Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes. *J. Mol. Evol.* 43, 484–516. doi: 10.1007/BF02337521
- Wang, R., Farrona, S., Vincent, C., Joecker, A., Schoof, H., Turck, F., et al. (2009). PEP1 regulates perennial flowering in *Arabidopsis*. *Nature* 459, 423–427. doi: 10.1038/nature07988
- Werner, J. D., Borevitz, J. O., Uhlenhaut, N. H., Ecker, J. R., Chory, J., and Weigel, D. (2005). FRIGIDA-independent variation in flowering time of natural *Arabidopsis thaliana* accessions. *Genetics* 170, 1197–1207. doi: 10.1534/genetics.104.036533
- Yang, L., Wang, H.-N., Hou, X.-H., Zou, Y.-P., Han, T.-S., Niu, X.-M., et al. (2018). Parallel evolution of common allelic variants confers flowering diversity in *Capsella rubella*. *Plant Cell* doi: 10.1105/tpc.18.00124 [Epub ahead of print].
- Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088
- Yanovsky, M. J., and Kay, S. A. (2002). Molecular basis of seasonal time measurement in *Arabidopsis*. *Nature* 419, 308–312. doi: 10.1038/nature00996
- Zhang, J., Nielsen, R., and Yang, Z. (2005). Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol.* 22, 2472–2479. doi: 10.1093/molbev/msi2372

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 © 2018 Lee, Hsieh, Schranz and Mitchell-Olds. 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.



How to Evolve a Perianth: A Review of Cadastral Mechanisms for Perianth Identity

Marie Monniaux* and Michiel Vandenbussche

Laboratoire Reproduction et Développement des Plantes, Université de Lyon, ENS de Lyon, UCB Lyon 1, CNRS, INRA, Lyon, France

OPEN ACCESS

Edited by:

Annette Becker,
Justus-Liebig-Universität Gießen,
Germany

Reviewed by:

Elena M. Kramer,
Harvard University, United States
Rainer Melzer,
University College Dublin, Ireland

*Correspondence:

Marie Monniaux
marie.monniaux@ens-lyon.fr

Specialty section:

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

Received: 18 June 2018

Accepted: 09 October 2018

Published: 29 October 2018

Citation:

Monniaux M and
Vandenbussche M (2018) How
to Evolve a Perianth: A Review
of Cadastral Mechanisms for Perianth
Identity. *Front. Plant Sci.* 9:1573.
doi: 10.3389/fpls.2018.01573

The flower of angiosperms is considered to be a major evolutionary innovation that impacted the whole biome. In particular, two properties of the flower are classically linked to its ecological success: bisexuality and a differentiated perianth with sepals and petals. Although the molecular basis for floral organ identity is well understood in extant species and summarized in the famous ABC model, how perianth identity appeared during evolution is still unknown. Here we propose that cadastral mechanisms that maintain reproductive organ identities to the center of the flower could have supported perianth evolution. In particular, repressing B- and C-class genes expression toward the inner whorls of the flower, is a key process to isolate domains with sepal and petal identity in the outer whorls. We review from the literature in model species the diverse regulators that repress B- and C-class genes expression to the center of the flower. This review highlights the existence of both unique and conserved repressors between species, and possible candidates to investigate further in order to shed light on perianth evolution.

Keywords: perianth, flower, evolution, petal, sepal, ABC model

INTRODUCTION

Flowering plants (angiosperms) gather more than 350,000 species, a stunning number in regard to all other land plants that count no more than 35,000 species (The Plant List, 2013). This dominance of angiosperms might be partly due to the flower, a highly efficient structure for reproduction (Regal, 1977). The flower has some key features such as bisexuality, a closed carpel, and a perianth (i.e., the structure that surrounds the reproductive organs, typically organized in sepals, and petals) that can attract pollinators and therefore participate in the speciation process (Fenster et al., 2004). This is mainly supported by the petals, that display a complex set of traits such as color, fragrance, shape, or epidermal cell patterns (Glover, 2014). Petals can also assist in flower opening (van Doorn and Van Meeteren, 2003), while sepals mainly protect the other floral organs. In this review we will use the term petal and sepal as a functional definition for all petaloid (showy and playing an attractive role) and sepaloid (greenish and playing a protective role) organs, respectively, irrespective of their position in the flower. With this definition, all petals (and all sepals) are therefore not necessarily homologous organs (Ronse De Craene and Brockington, 2013).

Although recent research has led to considerable progress on the question of the origin of the flower (Moyroud et al., 2017; Sauquet et al., 2017), large questions are still open.

In particular, the timing and order of events leading from the reproductive structure of the most recent common ancestor of seed plants – likely a unisexual structure without perianth – to the ancestral flower – likely a bisexual flower with an undifferentiated perianth of petals – is still unknown (Sauquet and Magallón, 2018). These events include transition from unisexuality to bisexuality, compression of the reproductive axis, evolution of a perianth and evolution of a closed carpel (Specht and Bartlett, 2009; Sauquet and Magallón, 2018; Scutt, 2018). Despite this uncertainty it seems reasonable to assume that the perianth evolved last, after bisexuality, axis compression and carpel evolution (Baum and Hileman, 2007). Later, the perianth often differentiated into an outer whorl of sepals and an inner whorl of petals (resulting in a so-called differentiated or bipartite perianth), which is particularly representative of core eudicots (Specht and Bartlett, 2009).

The origin of the perianth is still unresolved, but anatomical and developmental observations can shed some light on it. Sepals from most angiosperms have a leaf-like appearance suggesting they have a direct bract or leaf origin. Petals likely arose multiple times during evolution (Kramer and Irish, 2000) with two possible origins: bracteopetals that evolved from bracts and andropetals that evolved from stamens. Bracteopetals are typically observed in basal angiosperms that show a continuous differentiation between bracts and petaloid organs (Ronse De Craene, 2007). In contrast andropetals appear restricted to a few clades (Ranunculales and Caryophyllales for instance) where petals have probably been lost and reinvented (Ronse De Craene, 2007; Brockington et al., 2012; Ronse De Craene and Brockington, 2013). However, for most angiosperm species, the origin of petals remains unclear and a combination of anatomical and genetic work are needed to discriminate between the two possibilities.

Genetic work on model species have provided molecular support for the key events accompanying flowering: formation of the flower meristem, specification of floral organ identities (the famous ABC model), floral organ outgrowth and maturation, and fertilization (Glover, 2014). Based on this data, molecular models for the evolution of floral structures such as the bisexual axis have been proposed (Baum and Hileman, 2007; Frohlich and Chase, 2007; Specht and Bartlett, 2009). A similar approach can be followed to generate molecular hypotheses for the origin of the perianth; more specifically to speculate how an identity domain for the perianth could have emerged from an ancestral flower containing only reproductive organs. Here we propose that cadastral mechanisms maintaining reproductive identity to the center of the flower could have supported perianth appearance during evolution. Based on genetic work in model species, we review some of the molecular players underlying these cadastral mechanisms.

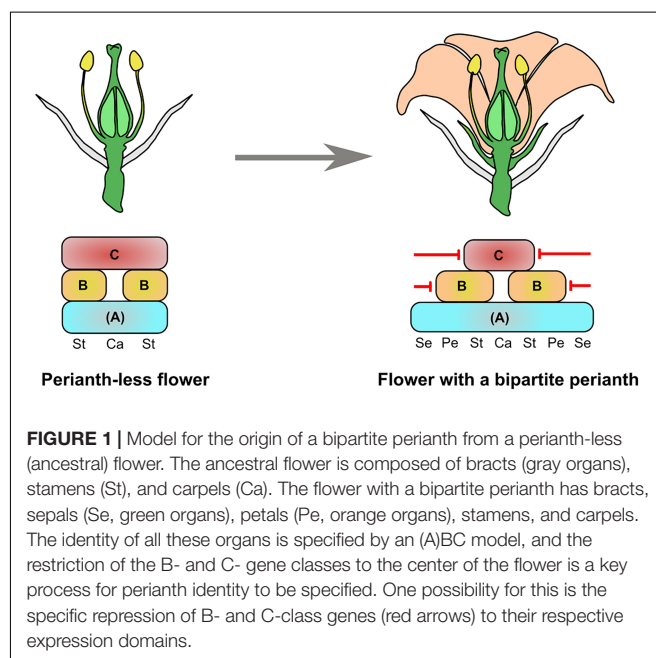
CREATING A DOMAIN FOR PERIANTH IDENTITY

Assuming that the perianth was the last angiosperm synapomorphy to appear, we have asked the following question

(**Figure 1**): how could a typical flower with 4 organ identities have been generated from a bisexual perianth-less flower with 2 organ identities?

In **Figure 1**, we use a genetic framework based on the (A)BC or FBC models proposed by Causier et al. or Baum and Hileman, respectively, where (A) or F is a floral identity function acquired at early stages of floral meristem development and necessary for all floral organ identities (Baum and Hileman, 2007; Causier et al., 2010). We chose not to use the classical ABC model because the existence of the A function is much debated (Schwarz-Sommer et al., 1990; Litt, 2007; Causier et al., 2010; Heijmans et al., 2012; Morel et al., 2017). With the (A)BC framework, floral organ identities from a typical angiosperm flower are specified by a combination of (A) expression for sepals, (A) + B for petals, (A) + B + C for stamens and (A) + C for carpels. Similarly, we can assume that the perianth-less ancestral flower had floral organ identities specified by (A) + B + C for stamens and (A) + C for carpels. In support of this, the B- and C-class functions are remarkably conserved across angiosperms, and gymnosperm male and female cones also show B + C and C gene expression, respectively (Sundström and Engström, 2002; Becker and Theissen, 2003; Zhang et al., 2004; Moyroud et al., 2017).

Therefore, the difference in gene expression between the perianth-less and bipartite perianth flower mainly resides in the peripheral expression domains of the B and C genes. This suggests that cadastral mechanisms that maintain (through repression or activation) B- and C-class gene expression to their dedicated area could have been key for perianth evolution. From reviewing the literature, it appears that several repressors of B- and C-class gene expression have been identified in extant species. Indeed, specifically repressing B genes from the first whorl and C genes from the first and second whorls is one possible way to generate a domain sufficient for sepal and petal



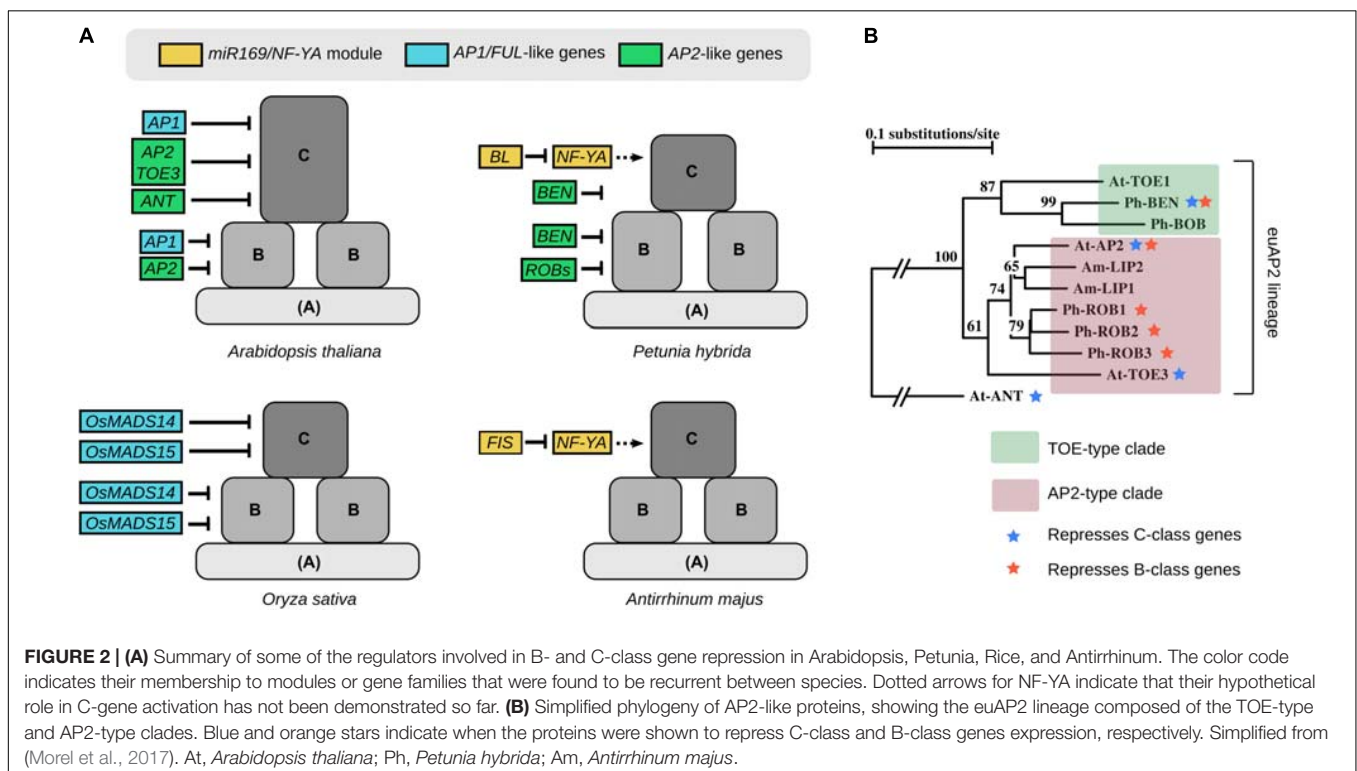
identity to be specified. In the following paragraphs we will review the repressors of B- and C-class genes and examine how conserved their function is across angiosperms (**Figure 2A**). This review is mostly based on functional studies in the model species *Arabidopsis thaliana* (Arabidopsis), *Petunia hybrida* (Petunia), *Oryza sativa* (Rice), and *Antirrhinum majus* (Antirrhinum).

REPRESSORS OF C-CLASS GENE EXPRESSION

A-class genes from the classical ABC model were proposed, already from the beginning, to have a dual function in specifying organ identity in the first two whorls (sepals and petals, alone or in combination with B-genes), and repressing the C function from these same whorls (Coen and Meyerowitz, 1991). In Arabidopsis, *APETALA1* (*AP1*) and *AP2* are generally classified as A-class genes, although *AP1* was added only later to this class (Bowman et al., 1991; Coen and Meyerowitz, 1991; Gustafson-Brown et al., 1994). Indeed, *AP1* does repress C-class gene expression from young flowers, by forming complexes with the flower meristem identity regulators *SHORT VEGETATIVE PHASE* (*SVP*) or *AGAMOUS-LIKE24* (*AGL24*) and the general floral repressors *LEUNIG* (*LUG*) and *SEUSS* (*SEU*) (Gregis et al., 2006, 2009; Sridhar et al., 2006). But examination of *ap1* in combination with other mutations revealed that *AP1* is not strictly necessary for petal and sepal identity, since these organs can develop in some *ap1* mutant backgrounds (Yu et al., 2004; Causier et al., 2010). Moreover outside of Brassicaceae, *AP1* orthologues are generally not needed for

sepal and petal identity (Litt, 2007; Causier et al., 2010). For instance in *Antirrhinum*, mutant in the *AP1* ortholog *SQUAMOSA* shows defects in floral meristem identity but not necessarily in perianth identity (Huijser et al., 1992). In contrast in Rice, mutations in the *AP1/FRUITFUL* (*FUL*) lineage members *OsMADS14* and *OsMADS15* results in the extension of C-class gene expression (and to a lesser extent, B-class gene expression) in the outer whorls of the flower, therefore leading to palea-to-carpel and lodicule-to-stamen homeotic conversions (Wu et al., 2017).

In Arabidopsis the other A-class gene *AP2* represses *AGAMOUS* (*AG*, the C-class gene) expression in whorls 1 and 2, and the *ap2* mutant shows the sepal-to-carpel and petal-to-stamen homeotic conversions expected from an A-class mutant (Drews et al., 1991) (see **Figure 2B** for a simplified phylogeny of *AP2*-like genes). Both the *AP2*-type gene *TARGET OF EAT 3* (*TOE3*) and the *AP2*-like gene *AINTEGUMENTA* (*ANT*) also repress *AG* expression; however, homeotic changes in the corresponding mutants are very subtle, if any (Krizek et al., 2000; Jung et al., 2014). *AP2* and *TOE3* expression is regulated at the translational level by the microRNA *miR172* (Chen, 2004; Wollmann et al., 2010). In Petunia, it was recently found that the *euAP2* clade member *BLIND ENHANCER* (*BEN*), although not the ortholog of *AP2*, also represses C-class gene expression from the first whorl (Morel et al., 2017); in maize mutations in the *AP2*-like genes *ids1* and *sid1* result in ectopic expression of the *AG*-like genes *zag1* and *zmm2* in bracts that become carpeloid (Chuck et al., 2008); in other species antagonistic expression patterns suggest a similar repressive role of *AP2*-like genes on C-class gene expression (Yang et al., 2015). In contrast, the



Antirrhinum AP2 orthologs *LIPLESS1* (*LIP1*) and *LIP2*, and the petunia AP2 orthologs *REPRESSOR OF B-FUNCTION 1* (*ROB1*), *ROB2*, and *ROB3* play a role in sepal and petal development but do not seem to antagonize C-class gene expression in the perianth (Keck et al., 2003; Morel et al., 2017). Overall, this shows that members of the *euAP2* lineage are often involved in C-class gene repression in the outer whorls of the flower, but this role sometimes has been swapped between members of the lineage, switching between AP2-type and TOE-type clade members (Figure 2B). These two clades predate the monocot-eudicot divergence (Kim et al., 2006), suggesting that the repression of C-class genes by members of the *euAP2* lineage might be relatively ancient.

In Arabidopsis, other C-class repressors have been identified: the SUPERMAN-like zinc finger protein RABBIT EARS (RBE) represses AG in whorl 2 (Krizek et al., 2006), while the bZIP transcription factor PERIANTHIA (PAN), together with SEU, represses AG in whorl 1 (Das et al., 2009; Wynn et al., 2014). However, mutations in these genes are not sufficient to cause homeotic changes in floral organs, showing that a certain threshold of AG ectopic expression is needed for homeotic conversion. In contrast, completely different C-class repressor genes have been identified in Petunia and Antirrhinum. Indeed in both species a member of the *miR169* family, *BLIND* (*BL*) in Petunia and *FISTULATA* (*FIS*) in Antirrhinum, represses C-class gene expression in the outer whorls, possibly by targeting members of the NF-YA family for degradation (Cartolano et al., 2007). Members of the *miR169*/NF-YA module from various species have been involved in several developmental processes such as flowering time, root architecture, embryogenesis, general responses to stress or interaction with pathogens (Laloum et al., 2013; Zhao et al., 2016; Zanetti et al., 2017) but apart from Petunia and Antirrhinum, never in floral patterning. This specific function of *miR169*/NF-YA might thus have evolved only in the euasterids lineage, unless in rosids it is hidden by redundancy between the multiple copies of *miR169* and NF-YA genes.

Other C-class repressors with a broader spatial action have been found, including BELLRINGER that represses AG expression in the Arabidopsis stem, inflorescence meristem and young flower meristem (Bao et al., 2004); CURLY LEAF that represses AG expression in all vegetative tissues of Arabidopsis and Brachypodium (Goodrich et al., 1997; Lomax et al., 2018); STERILE APETALA that represses AG expression in the outer floral whorls and in the inflorescence meristem of Arabidopsis (Byzova et al., 1999); and FILAMENTOUS FLOWER that represses AG expression in the outer floral whorls and in the peduncle of Arabidopsis (Chen et al., 1999). Mutations in each of these genes cause homeotic defects in Arabidopsis floral organs. The fact that these repressors are not spatially specific to the outer floral whorls does not exclude that they could have played a role in perianth evolution by being coopted for repression of AG in whorls 1 and whorls 2, while they already repressed AG expression from other tissues (True and Carroll, 2002). However, since their role has hardly been investigated in other species than Arabidopsis so far, we do not know how conserved these mechanisms are and if they could have been involved in perianth evolution.

REPRESSORS OF B-CLASS GENE EXPRESSION

In contrast to the many C-class genes repressor identified, fewer genes have been shown to repress B-class gene expression from the first whorl of the flower (Figure 2). In Arabidopsis, AP2 represses the B-class genes *APETALA3* (*AP3*) and *PISTILLATA* (*PI*) expression in whorl 1. While this is not evident from the phenotype of single *ap2* mutants, this becomes visible when an *ap2* mutant allele is present in an heterozygous state in the *topless* (*tpl*) mutant background, resulting in a partial sepal-to-petal conversion due to ectopic *PI* and *AP3* expression in whorl 1 (Krogan et al., 2012). But by far, the clearest evidence of B-class gene derepression in whorl 1 is found in Petunia in the quadruple *ben rob1 rob2 rob3* mutant, that shows an almost perfect homeotic conversion of sepals into petals (Morel et al., 2017). As such this flower is reminiscent of the undifferentiated perianth found in many angiosperm species, such as tulip or magnolia, and likely a characteristic trait of the ancestral flower (Sauquet et al., 2017). As previously mentioned, *BEN* and *ROBs* are all members of the *euAP2* lineage (Figure 2B). The single *ben* mutant also shows some petaloid sectors in sepals, indicating that *BEN* and *ROBs* partially redundantly repress B-genes expression in whorl 1 (Morel et al., 2017). Therefore, repression of B-genes expression by *euAP2* genes appears conserved between Arabidopsis and Petunia, suggesting that this regulation might have originated prior to the rosids-asterids divergence.

APETALA1 has also been proposed to repress *PI* and *AP3* expression in Arabidopsis, when part of the repressive complex with AGL24, SVP, LUG, and SEU (Gregis et al., 2006, 2009), but whether AP1 is directly involved in this regulation is unclear. In Rice the *osmads14 osmads15* double mutant shows some derepression of B-class gene expression in whorl 1 but this ectopic expression does not seem strong enough to alter organ identity (Wu et al., 2017). Altogether, this suggests a somehow conserved role of *AP1/FUL* genes in repressing B-class gene expression but the evidence is scarcer than for *AP2* genes.

CONSERVED AND UNIQUE REPRESSORS OF B- AND C-CLASS GENES

Our review highlights that members of the large AP2 family, and in particular the *euAP2* lineage, might have a conserved function in repressing both B- and C-class gene expression from the outer whorls of the flower, predating the rosids-asterids divergence. Hence members from this family are possible candidates to have played a role in perianth evolution. However, if *euAP2* genes were involved in both sepal and petal evolution, it would require uncoupling of their repressive action on B- and C-class genes, since B-class genes should be repressed in the first whorl only while C-class genes should be repressed in the first two whorls. In Arabidopsis, it is unknown how AP2 can have whorl-specific repressive action on B- and C-class genes, but it possibly resides in the interaction with different protein partners between whorls.

In *Petunia*, euAP2 proteins repress B- and C-gene expression in the first whorl only, while repression of C-class genes in the second whorl is completed by *BL*, showing how the dual repressive function has been distributed between two sets of genes in this species. More functional studies in angiosperms, and particularly in early diverging taxa, are now needed to evaluate the possibility that *euAP2* genes were involved in perianth evolution.

Our review also shows that a large variety of repressors of C- (and to a more minor extent, B-) class genes exist in extant species. One might wonder in particular why so many C-genes repressors are found. None of the identified repressors are redundant with each other since single mutants in question all show ectopic C expression. Hence instead of conferring robust repression of C-class genes expression, these numerous repressors might actually provide a multiplicity of ways for evolution to relieve C-expression in the perianth. Since the perianth is a highly evolvable structure, i.e., a flexible trait that evolved multiple times during angiosperm evolution (Baum and Whitlock, 1999; Kramer and Irish, 2000; Hileman and Irish, 2009; Geuten et al., 2011), an hypothesis is that evolution could have tinkered with these various repressors for the perianth to appear or disappear in different taxa.

Petaloid features are not exclusively found on second whorl petals but often have been transferred to sepals, bracts or even stamens. Flower morphology is remarkably flexible and while true petals may have been lost, petaloidy shifted to analogous organs that acquired the petal traits needed for recognition by pollinators. These petaloid features are sometimes correlated with ectopic B-gene expression, like in many non-grass monocots for instance (Kanno et al., 2007), but there are also many cases of petaloid structures that have little or no B-genes expression, as well as non-petaloid structures that do express B-class genes, as reviewed in (Ronse De Craene and Brockington, 2013). We can argue that not all B-class genes might have been identified in species without a sequenced genome. Still, the evolution of

petaloidy appears to be more complex than mere shifts in B-class gene expression, and the genes underlying these transfers of function from one organ to another might still remain to be identified.

CONCLUSION

In this review we proposed that repressing reproductive organ identity to the center of the flower is a possible way for perianth identity to have emerged in the periphery of the flower, and we reviewed the B- and C-class genes repressors that have been identified in model species. Whether these repressors were actually evolved in perianth evolution some 150 million years ago remains of course hypothetical, and functional experiments in early-diverging gymnosperms and angiosperms are needed to evaluate such hypotheses. By far the largest source of variation in perianth morphology in angiosperms does not lie in flower patterning, but in changes in shape, color, scent or size of the petals, as beautifully illustrated in (Byng et al., 2018). The genetic basis for these variations, some of quantitative nature like spur length in *Aquilegia* (Yant et al., 2015), others of qualitative nature like presence or absence of pigmentation in *Petunia* (Hoballah et al., 2007), has only been identified on few occasions (Moyroud and Glover, 2017). These genes are likely downstream targets of B-class regulators, but how these master developmental genes direct the establishment of all petaloid features in a simultaneous manner, and which parts of this large network have been modified during evolution to generate morphological diversity, remains a mystery as big as perianth evolution itself.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

- Bao, X., Franks, R. G., Levin, J. Z., and Liu, Z. (2004). Repression of AGAMOUS by BELLRINGER in floral and inflorescence meristems. *Plant Cell* 16, 1478–1489. doi: 10.1105/tpc.021147
- Baum, D. A., and Hileman, L. C. (2007). “A developmental genetic model for the origin of the flower,” in *Annual Plant Reviews: Flowering and its Manipulation*, Vol. 20, ed. C. Ainsworth (Chichester: Blackwell Publishing Ltd), 1–27. doi: 10.1002/9780470988602.ch1
- Baum, D. A., and Whitlock, B. A. (1999). Plant development: genetic clues to petal evolution. *Curr. Biol.* 9, R525–R527. doi: 10.1016/S0960-9822(99)80327-3
- Becker, A., and Theissen, G. (2003). The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Mol. Phylogenet. Evol.* 29, 464–489. doi: 10.1016/S1055-7903(03)00207-0
- Bowman, J. L., Smyth, D. R., and Meyerowitz, E. M. (1991). Genetic interactions among floral homeotic genes of *Arabidopsis*. *Development* 112, 1–20.
- Brockington, S. F., Rudall, P. J., Frohlich, M. W., Oppenheimer, D. G., Soltis, P. S., and Soltis, D. E. (2012). ‘Living stones’ reveal alternative petal identity programs within the core eudicots. *Plant J.* 69, 193–203. doi: 10.1111/j.1365-313X.2011.04797.x
- Byng, J. W., Smets, E. F., van Vugt, R., Bidault, E., Davidson, C., Kenicer, G., et al. (2018). *Angiosperm Poster. Plant Gateway*. Available at: <http://www.plantgateway.com/poster/> [accessed April 25, 2018].
- Byzova, M. V., Franken, J., Aarts, M. G., de Almeida-Engler, J., Engler, G., Mariani, C., et al. (1999). *Arabidopsis* STERILE APETALA, a multifunctional gene regulating inflorescence, flower, and ovule development. *Genes Dev.* 13, 1002–1014. doi: 10.1101/gad.13.8.1002
- Cartolano, M., Castillo, R., Efremova, N., Kuckenberg, M., Zethof, J., Gerats, T., et al. (2007). A conserved microRNA module exerts homeotic control over *Petunia hybrida* and *Antirrhinum majus* floral organ identity. *Nat. Genet.* 39, 901–905. doi: 10.1038/ng2056
- Causier, B., Schwarz-Sommer, Z., and Davies, B. (2010). Floral organ identity: 20 years of ABCs. *Semin. Cell Dev. Biol.* 21, 73–79. doi: 10.1016/j.semcdb.2009.10.005
- Chen, Q., Atkinson, A., Otsuga, D., Christensen, T., Reynolds, L., and Drews, G. N. (1999). The *Arabidopsis* FILAMENTOUS FLOWER gene is required for flower formation. *Development* 126, 2715–2726.
- Chen, X. (2004). A MicroRNA as a translational repressor of *APETALA2* in *Arabidopsis* flower development. *Science* 303, 2022–2025. doi: 10.1126/science.1088060

- Chuck, G., Meeley, R., and Hake, S. (2008). Floral meristem initiation and meristem cell fate are regulated by the maize *AP2* genes *ids1* and *sid1*. *Development* 135, 3013–3019. doi: 10.1242/dev.024273
- Coen, E. S., and Meyerowitz, E. M. (1991). The war of the whorls: genetic interactions controlling flower development. *Nature* 353, 31–37. doi: 10.1038/353031a0
- Das, P., Ito, T., Wellmer, F., Vernoux, T., Dedieu, A., Traas, J., et al. (2009). Floral stem cell termination involves the direct regulation of *AGAMOUS* by *PERANTHIA*. *Development* 136, 1605–1611. doi: 10.1242/dev.035436
- Drews, G. N., Bowman, J. L., and Meyerowitz, E. M. (1991). Negative regulation of the Arabidopsis homeotic gene *AGAMOUS* by the *APETALA2* product. *Cell* 65, 991–1002. doi: 10.1016/0092-8674(91)90551-9
- Fenster, C. B., Armbruster, W. S., Wilson, P., Dudash, M. R., and Thomson, J. D. (2004). Pollination syndromes and floral specialization. *Annu. Rev. Ecol. Evol. Syst.* 35, 375–403. doi: 10.1146/annurev.ecolsys.34.011802.132347
- Frohlich, M. W., and Chase, M. W. (2007). After a dozen years of progress the origin of angiosperms is still a great mystery. *Nature* 450, 1184–1189. doi: 10.1038/nature06393
- Geuten, K., Viaene, T., and Irish, V. F. (2011). Robustness and evolvability in the B-system of flower development. *Ann. Bot.* 107, 1545–1556. doi: 10.1093/aob/mcr061
- Glover, B. (2014). *Understanding Flowers and Flowering*, 2nd Edn. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199661596.001.0001
- Goodrich, J., Puangsomlee, P., Martin, M., Long, D., Meyerowitz, E. M., and Coupland, G. (1997). A Polycomb-group gene regulates homeotic gene expression in *Arabidopsis*. *Nature* 386, 44–51. doi: 10.1038/386044a0
- Gregis, V., Sessa, A., Colombo, L., and Kater, M. M. (2006). *AGL24*, *SHORT VEGETATIVE PHASE*, and *APETALA1* redundantly control *AGAMOUS* during early stages of flower development in *Arabidopsis*. *Plant Cell* 18, 1373–1382. doi: 10.1105/tpc.106.041798
- Gregis, V., Sessa, A., Dorca-Fornell, C., and Kater, M. M. (2009). The Arabidopsis floral meristem identity genes *AP1*, *AGL24* and *SVP* directly repress class B and C floral homeotic genes. *Plant J.* 60, 626–637. doi: 10.1111/j.1365-313X.2009.03985.x
- Gustafson-Brown, C., Savidge, B., and Yanofsky, M. F. (1994). Regulation of the Arabidopsis floral homeotic gene *APETALA1*. *Cell* 76, 131–143. doi: 10.1016/0092-8674(94)90178-3
- Heijmans, K., Morel, P., and Vandenbussche, M. (2012). MADS-box genes and floral development: the dark side. *J. Exp. Bot.* 63, 5397–5404. doi: 10.1093/jxb/ers233
- Hileman, L. C., and Irish, V. F. (2009). More is better: the uses of developmental genetic data to reconstruct perianth evolution. *Am. J. Bot.* 96, 83–95. doi: 10.3732/ajb.0800066
- Hoballah, M. E., Gubitz, T., Stuurman, J., Broger, L., Barone, M., Mandel, T., et al. (2007). Single gene-mediated shift in pollinator attraction in *Petunia*. *Plant Cell* 19, 779–790. doi: 10.1105/tpc.106.048694
- Huijser, P., Klein, J., Lonnig, W. E., Meijer, H., Saedler, H., and Sommer, H. (1992). Bracteomania, an inflorescence anomaly, is caused by the loss of function of the MADS-box gene *squamosa* in *Antirrhinum majus*. *EMBO J.* 11, 1239–1249. doi: 10.1002/j.1460-2075.1992.tb05168.x
- Jung, J.-H., Lee, S., Yun, J., Lee, M., and Park, C.-M. (2014). The miR172 target *TOE3* represses *AGAMOUS* expression during *Arabidopsis* floral patterning. *Plant Sci.* 21, 29–38. doi: 10.1016/j.plantsci.2013.10.010
- Kanno, A., Nakada, M., Akita, Y., and Hirai, M. (2007). Class B gene expression and the modified ABC model in nongrass monocots. *Sci. World J.* 7, 268–279. doi: 10.1100/tsw.2007.86
- Keck, E., McSteen, P., Carpenter, R., and Coen, E. (2003). Separation of genetic functions controlling organ identity in flowers. *EMBO J.* 22, 1058–1066. doi: 10.1093/emboj/cdg097
- Kim, S., Soltis, P. S., Wall, K., and Soltis, D. E. (2006). Phylogeny and domain evolution in the *APETALA2*-like gene family. *Mol. Biol. Evol.* 23, 107–120. doi: 10.1093/molbev/msj014
- Kramer, E. M., and Irish, V. F. (2000). Evolution of the petal and stamen developmental programs: evidence from comparative studies of the lower eudicots and basal angiosperms. *Int. J. Plant Sci.* 161, S29–S40. doi: 10.1086/317576
- Krizek, B. A., Lewis, M. W., and Fletcher, J. C. (2006). *RABBIT EARS* is a second-whorl repressor of *AGAMOUS* that maintains spatial boundaries in Arabidopsis flowers. *Plant J.* 45, 369–383. doi: 10.1111/j.1365-313X.2005.02633.x
- Krizek, B. A., Prost, V., and Macias, A. (2000). *AINTEGUMENTA* promotes petal identity and acts as a negative regulator of *AGAMOUS*. *Plant Cell* 12, 1357–1366. doi: 10.1105/tpc.12.8.1357
- Krogan, N. T., Hogan, K., and Long, J. A. (2012). *APETALA2* negatively regulates multiple floral organ identity genes in *Arabidopsis* by recruiting the co-repressor *TOPLESS* and the histone deacetylase *HDA19*. *Development* 139, 4180–4190. doi: 10.1242/dev.085407
- Laloum, T., De Mita, S., Gamas, P., Baudin, M., and Niebel, A. (2013). CCAAT-box binding transcription factors in plants: Y so many? *Trends Plant Sci.* 18, 157–166. doi: 10.1016/j.tplants.2012.07.004
- Litt, A. (2007). An evaluation of A-function: evidence from the *APETALA1* and *APETALA2* gene lineages. *Int. J. Plant Sci.* 168, 73–91. doi: 10.1086/509662
- Lomax, A., Woods, D. P., Dong, Y., Bouché, F., Rong, Y., Mayer, K. S., et al. (2018). An ortholog of *CURLY LEAF/ENHANCER OF ZESTE like-1* is required for proper flowering in *Brachypodium distachyon*. *Plant J.* 93, 871–882. doi: 10.1111/tpj.13815
- Morel, P., Heijmans, K., Rozier, F., Zethof, J., Chamot, S., Bento, S. R., et al. (2017). Divergence of the floral A-function between an asterid and a Rosid Species. *Plant Cell* 29, 1605–1621. doi: 10.1105/tpc.17.00098
- Moyroud, E., and Glover, B. J. (2017). The evolution of diverse floral morphologies. *Curr. Biol.* 27, R941–R951. doi: 10.1016/j.cub.2017.06.053
- Moyroud, E., Monniaux, M., Thévenon, E., Dumas, R., Scutt, C. P., Frohlich, M. W., et al. (2017). A link between *LEAFY* and B-gene homologues in *Welwitschia mirabilis* sheds light on ancestral mechanisms prefiguring floral development. *New Phytol.* 216, 469–481. doi: 10.1111/nph.14483
- Regal, P. J. (1977). Ecology and evolution of flowering plant dominance. *Science* 196, 622–629. doi: 10.1126/science.196.4290.622
- Ronse De Craene, L. P., and Brockington, S. F. (2013). Origin and evolution of petals in angiosperms. *Plant Ecol. Evol.* 146, 5–25. doi: 10.5091/plecevo.20.13.738
- Ronse De Craene, R. D. (2007). Are petals sterile stamens or bracts? The origin and evolution of petals in the core eudicots. *Ann. Bot.* 100, 621–630. doi: 10.1093/aob/mcm076
- Sauquet, H., and Magallón, S. (2018). Key questions and challenges in angiosperm macroevolution. *New Phytol.* 219, 1170–1187. doi: 10.1111/nph.15104
- Sauquet, H., von Balthazar, M., Magallón, S., Doyle, J. A., Endress, P. K., Bailes, E. J., et al. (2017). The ancestral flower of angiosperms and its early diversification. *Nat. Commun.* 8:16047. doi: 10.1038/ncomms16047
- Schwarz-Sommer, Z., Huijser, P., Nacken, W., Saedler, H., and Sommer, H. (1990). Genetic control of flower development by homeotic genes in *Antirrhinum majus*. *Science* 250, 931–936. doi: 10.1126/science.250.4983.931
- Scutt, C. P. (2018). “The origin of angiosperms,” in *Evolutionary Developmental Biology*, eds L. Nuno de la Rosa and G. Müller (Cham: Springer International Publishing), 1–20. doi: 10.1007/978-3-319-33038-9_60-1
- Specht, C. D., and Bartlett, M. E. (2009). Flower evolution: the origin and subsequent diversification of the angiosperm flower. *Annu. Rev. Ecol. Evol. Syst.* 40, 217–243. doi: 10.1146/annurev.ecolsys.110308.120203
- Sridhar, V. V., Surendrarao, A., and Liu, Z. (2006). *APETALA1* and *SEPALLATA3* interact with *SEUSS* to mediate transcription repression during flower development. *Development* 133, 3159–3166. doi: 10.1242/dev.02498
- Sundström, J., and Engström, P. (2002). Conifer reproductive development involves B-type MADS-box genes with distinct and different activities in male organ primordia. *Plant J.* 31, 161–169. doi: 10.1046/j.1365-313X.2002.01343.x
- The Plant List (2013). Available at: <http://www.theplantlist.org/>
- True, J. R., and Carroll, S. B. (2002). Gene co-option in physiological and morphological evolution. *Annu. Rev. Cell Dev. Biol.* 18, 53–80. doi: 10.1146/annurev.cellbio.18.020402.140619
- van Doorn, W. G., and Van Meeteren, U. (2003). Flower opening and closure: a review. *J. Exp. Bot.* 54, 1801–1812. doi: 10.1093/jxb/erg213
- Wollmann, H., Mica, E., Todesco, M., Long, J. A., and Weigel, D. (2010). On reconciling the interactions between *APETALA2*, *miR172* and *AGAMOUS* with the ABC model of flower development. *Development* 137, 3633–3642. doi: 10.1242/dev.036673
- Wu, F., Shi, X., Lin, X., Liu, Y., Chong, K., Theissen, G., et al. (2017). The ABCs of flower development: mutational analysis of *AP1/FUL*-like genes in rice

- provides evidence for a homeotic (A)-function in grasses. *Plant J.* 89, 310–324. doi: 10.1111/tpj.13386
- Wynn, A. N., Seaman, A. A., Jones, A. L., and Franks, R. G. (2014). Novel functional roles for PERIANTHIA and SEUSS during floral organ identity specification, floral meristem termination, and gynoecial development. *Front. Plant Sci.* 5:130. doi: 10.3389/fpls.2014.00130
- Yang, F. X., Zhu, G. F., Wang, Z., Liu, H. L., and Huang, D. (2015). A putative miR172-targeted *CeAPETALA2-like* gene is involved in floral patterning regulation of the orchid *Cymbidium ensifolium*. *Genet. Mol. Res.* 14, 12049–12061. doi: 10.4238/2015.October.5.18
- Yant, L., Collani, S., Puzey, J., Levy, C., and Kramer, E. M. (2015). Molecular basis for three-dimensional elaboration of the *Aquilegia* petal spur. *Proc. Biol. Sci.* 282:20142778. doi: 10.1098/rspb.2014.2778
- Yu, H., Ito, T., Wellmer, F., and Meyerowitz, E. M. (2004). Repression of AGAMOUS-LIKE 24 is a crucial step in promoting flower development. *Nat. Genet.* 36, 157–161. doi: 10.1038/ng1286
- Zanetti, M. E., Ripodas, C., and Niebel, A. (2017). Plant NF-Y transcription factors: key players in plant-microbe interactions, root development and adaptation to stress. *Biochim. Biophys. Acta* 1860, 645–654. doi: 10.1016/j.bbagr.2016.11.007
- Zhang, P., Tan, H. T. W., Pwee, K.-H., and Kumar, P. P. (2004). Conservation of class C function of floral organ development during 300 million years of evolution from gymnosperms to angiosperms. *Plant J.* 37, 566–577. doi: 10.1046/j.1365-313X.2003.01983.x
- Zhao, H., Wu, D., Kong, F., Lin, K., Zhang, H., and Li, G. (2016). The *Arabidopsis thaliana* nuclear factor Y transcription factors. *Front. Plant Sci.* 7:2045. doi: 10.3389/fpls.2016.02045

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 © 2018 Monniaux and Vandenbussche. 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.



Gene Duplication and Transference of Function in the paleoAP3 Lineage of Floral Organ Identity Genes

Kelsey D. Galimba[†], Jesús Martínez-Gómez and Verónica S. Di Stilio^{*}

Department of Biology, University of Washington, Seattle, WA, United States

OPEN ACCESS

Edited by:

Elena M. Kramer,
Harvard University, United States

Reviewed by:

Pablo Daniel Jenik,
Franklin & Marshall College,
United States
Madelaine Elisabeth Bartlett,
University of Massachusetts Amherst,
United States

*Correspondence:

Verónica S. Di Stilio
distilio@u.washington.edu

[†]Present address:

Kelsey D. Galimba,
Appalachian Fruit Research Station,
United States Department of
Agriculture – Agricultural Research
Service, Kearneysville, WV,
United States

Specialty section:

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

Received: 08 November 2017

Accepted: 28 February 2018

Published: 23 March 2018

Citation:

Galimba KD, Martínez-Gómez J and
Di Stilio VS (2018) Gene Duplication
and Transference of Function
in the paleoAP3 Lineage of Floral
Organ Identity Genes.
Front. Plant Sci. 9:334.
doi: 10.3389/fpls.2018.00334

The floral organ identity gene *APETALA3* (*AP3*) is a MADS-box transcription factor involved in stamen and petal identity that belongs to the B-class of the ABC model of flower development. *Thalictrum* (Ranunculaceae), an emerging model in the non-core eudicots, has *AP3* homologs derived from both ancient and recent gene duplications. Prior work has shown that petals have been lost repeatedly and independently in Ranunculaceae in correlation with the loss of a specific *AP3* paralog, and *Thalictrum* represents one of these instances. The main goal of this study was to conduct a functional analysis of the three *AP3* orthologs present in *Thalictrum thalictroides*, representing the paleo*AP3* gene lineage, to determine the degree of redundancy versus divergence after gene duplication. Because *Thalictrum* lacks petals, and has lost the petal-specific *AP3*, we also asked whether heterotopic expression of the remaining *AP3* genes contributes to the partial transference of petal function to the first whorl found in insect-pollinated species. To address these questions, we undertook functional characterization by virus-induced gene silencing (VIGS), protein–protein interaction and binding site analyses. Our results illustrate partial redundancy among *Thalictrum AP3*s, with deep conservation of B-class function in stamen identity and a novel role in ectopic petaloidy of sepals. Certain aspects of petal function of the lost *AP3* locus have apparently been transferred to the other paralogs. A novel result is that the protein products interact not only with each other, but also as homodimers. Evidence presented here also suggests that expression of the different *ThtAP3* paralogs is tightly integrated, with an apparent disruption of B function homeostasis upon silencing of one of the paralogs that codes for a truncated protein. To explain this result, we propose two testable alternative scenarios: that the truncated protein is a dominant negative mutant or that there is a compensational response as part of a back-up circuit. The evidence for promiscuous protein–protein interactions via yeast two-hybrid combined with the detection of *AP3* specific binding motifs in all B-class gene promoters provide partial support for these hypotheses.

Keywords: ABC model, MADS-box genes, B-class genes, ectopic petaloidy, flower development, non-core eudicot, VIGS, *Thalictrum thalictroides*

INTRODUCTION

Gene duplication has long been interpreted as a potential source of raw genetic material acted upon by evolution (Ohno, 1970; Soukup, 1974; Force et al., 1999). Genes in the ABC model of flower development have been an especially targeted, undergoing multiple duplication events during the course of angiosperm evolution (Theissen et al., 1996; Airoidi and Davies, 2012). Because changes to floral organ identity genes have a profound effect on flower development, they present an ideal system to study the outcome of gene duplication in relation to morphological adaptation (Soltis and Soltis, 2014).

In the model angiosperm *Arabidopsis thaliana*, the B-class consists of two members, *APETALA3* (AP3) and *PISTILLATA* (PI). Both are necessary in combination with the E-class genes *SEPALLATA1–4* (*SEP1–4*), for petal identity in the second and stamen identity in the third whorl (Bowman et al., 1989, 1991; Coen and Meyerowitz, 1991; Theissen and Saedler, 2001). When B-class function is lost, stamen primordia are homeotically converted into carpels and petal primordia into sepals (Krizek and Meyerowitz, 1996). B-class gene function is highly conserved throughout angiosperms, occurring in core and non-core eudicots as well as monocots (Kim et al., 2004; Zahn et al., 2005; Litt and Kramer, 2010; Di Stilio, 2011). Nevertheless, there are examples of B-class genes being expressed in other plant organs, e.g., in root nodules of alfalfa (Heard and Dunn, 1995) and in first whorl petaloid tepals of tulips (Kanno et al., 2003), suggesting that they may have adopted novel roles in several lineages.

After an ancient duplication leading to the AP3 and PI lineages, in the stem group of the order Ranunculales, two duplication events led to three paralogous lineages of AP3: AP3-I, AP3-II, and AP3-III (Kramer et al., 2003; Figure 1A). *Thalictrum thalictroides* (Ranunculaceae), a member of the Ranunculid lineage that is sister to all other eudicots (Soltis et al., 2011; Maia et al., 2014), has one PI (*ThtPI*) and three AP3 orthologs (*ThtAP3-1*, *ThtAP3-2a*, and *ThtAP3-2b*) (Kramer et al., 2003; Di Stilio et al., 2005). The paralog *ThtAP3-1* belongs to the AP3-I clade, while *ThtAP3-2a* and *ThtAP3-2b* belong to the AP3-II clade and are products of a more recent duplication that likely occurred in the common ancestor of *Thalictrum* and *Aquilegia* (Sharma et al., 2011). *ThtAP3-2a* contains a premature stop, resulting in a truncation affecting the conserved C-terminal motif of the protein; 18 residues encompassing half of the “PI motif-derived” and all the “paleoAP3” motif is missing (Figure 1B; Di Stilio et al., 2005). No AP3-III paralog has been identified so far in *Thalictrum* by PCR (Di Stilio et al., 2005; Zhang et al., 2013) or by BLAST search of available transcriptomes (unpublished data). Functional studies of the *Aquilegia coerulea* AP3-III ortholog provide evidence for sub-functionalization of this gene to petal identity (Sharma et al., 2011) and this function appears to be conserved throughout the order, including Papaveraceae (Arango-Ocampo et al., 2016); the lack of an AP3-III ortholog in *Thalictrum* therefore correlates with the loss of petals in the genus (Di Stilio et al., 2005; Zhang et al., 2013).

Petals may be interpreted as adaptive structures that attract biotic pollinators and this function has been transferred,

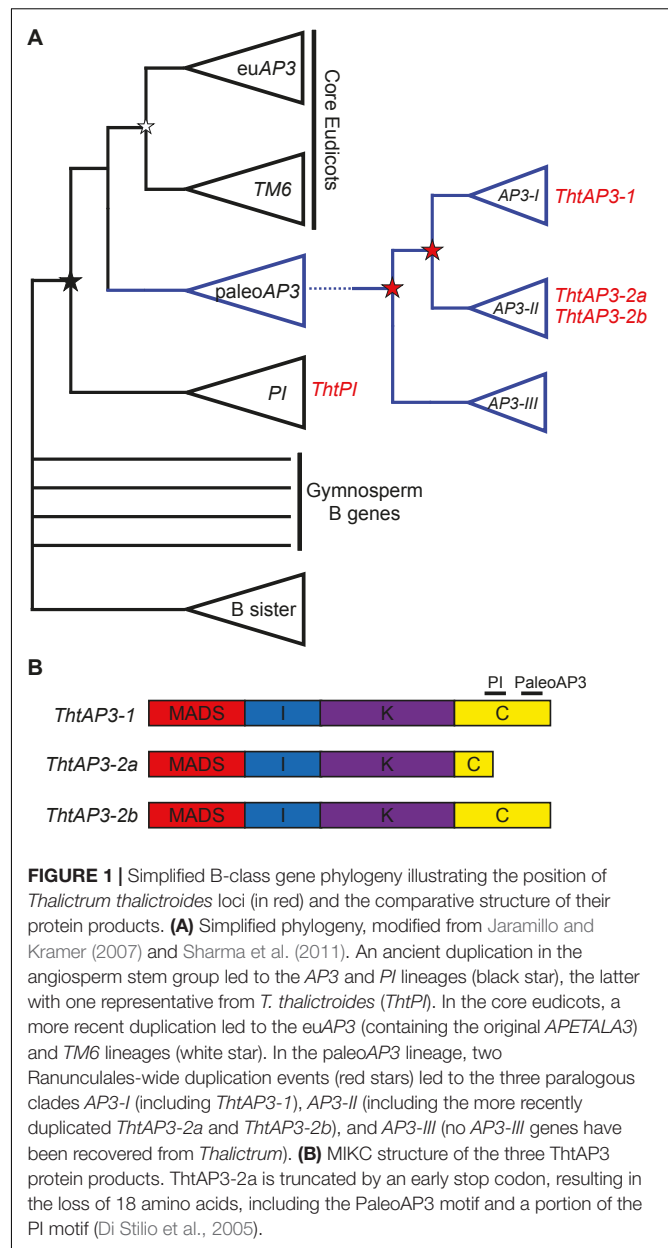


FIGURE 1 | Simplified B-class gene phylogeny illustrating the position of *Thalictrum thalictroides* loci (in red) and the comparative structure of their protein products. **(A)** Simplified phylogeny, modified from Jaramillo and Kramer (2007) and Sharma et al. (2011). An ancient duplication in the angiosperm stem group led to the AP3 and PI lineages (black star), the latter with one representative from *T. thalictroides* (*ThtPI*). In the core eudicots, a more recent duplication led to the euAP3 (containing the original *APETALA3*) and TM6 lineages (white star). In the paleoAP3 lineage, two Ranunculales-wide duplication events (red stars) led to the three paralogous clades AP3-I (including *ThtAP3-1*), AP3-II (including the more recently duplicated *ThtAP3-2a* and *ThtAP3-2b*), and AP3-III (no AP3-III genes have been recovered from *Thalictrum*). **(B)** MIKC structure of the three ThtAP3 protein products. *ThtAP3-2a* is truncated by an early stop codon, resulting in the loss of 18 amino acids, including the PaleoAP3 motif and a portion of the PI motif (Di Stilio et al., 2005).

fully or in part, to other organ types multiple times during angiosperm evolution (Cronk et al., 2002). In fact, in the Ranunculaceae, where sepals are often petaloid, petals have been lost independently multiple times with the role of attraction adopted instead by showy sepals (Zhang et al., 2013). In a number of distantly related species, including *Aristolochia* and *Tulipa*, heterotopic expression of B-class genes has been proposed as causative of petaloidy of first-whorl organs (van Tunen et al., 1993; Kanno et al., 2003; Jaramillo and Kramer, 2004). In *Aquilegia*, *AqAP3s* are not involved in petal identity or development of papillate epidermal cells, although they do contribute to the production of anthocyanin in petaloid sepals (Jaramillo and Kramer, 2007; Sharma et al., 2011; Sharma and Kramer, 2017). Although *T. thalictroides* flowers lack petals

from inception, its sepals have a number of features typically associated with petals: they are non-photosynthetic (white or pink), relatively large (surpassing stamens), they may contain papillate cells in the upper epidermis (Di Stilio et al., 2009) and they express all B-class genes (Di Stilio et al., 2005). Sepals of wind pollinated species such as *Thalictrum dioicum*, on the other hand, are smaller (shorter than stamens), green, lack papillate cells, and do not express AP3 orthologs (Di Stilio et al., 2005; Soza et al., 2012).

The goal of this study was to conduct a functional analysis of the three AP3 orthologs in *T. thalictroides*, a ranunculid with representatives of the paleoAP3 lineage of B-class genes. We aimed to determine the degree of redundancy versus divergence amongst these paralogs, while comparing their role to the more widely characterized euAP3 lineage. Finally, we investigated whether one or more of these loci are involved in ectopic petaloidy, in the form of partial transference of petal function to sepals, in the insect-pollinated species *T. thalictroides*. We addressed these questions via targeted silencing of individual genes, by determining the ability of each protein product to dimerize and by identifying AP3 binding sites within promoter regions of all four B-class genes.

MATERIALS AND METHODS

Gene Sequences

Partial coding sequences for *ThtAP3-1*, *ThtAP3-2a*, and *ThtAP3-2b* were obtained from GenBank (AY162886, AY162887, and AY162888; Kramer et al., 2003). Full coding sequences were obtained from a 1KP *T. thalictroides* transcriptome¹. Full genomic sequences and promoter regions were obtained from a *T. thalictroides* reference genome (unpublished) and deposited in GenBank (MG889397, MG889396, and MG889395).

Plant Materials

Thalictrum thalictroides bare root plants were purchased from nurseries and grown in the University of Washington (UW) Greenhouse under ambient conditions from mid-February through mid-May. A voucher specimen is deposited at the UW Herbarium (V. Di Stilio 123, WTU 376542).

Virus-Induced Gene Silencing

Regions for targeted gene silencing of *Thalictrum* AP3 genes were selected to exclude areas with more than 13 continuous homologous base pairs among the three paralogs (gene duplicates). Fragments were amplified by PCR from a representative clone using primers with added *Bam*HI and *Kpn*I restriction sites (Supplementary Table S1), ligated to linearized tobacco rattle virus vector (TRV2; Liu et al., 2002) and transformed into *Agrobacterium tumefaciens* strain GV3101. The TRV2-*ThtAP3-1* construct was prepared using a 429 bp fragment comprising the C-terminal region (282 bp) and 3'UTR (147 bp) of *ThtAP3-1*. The TRV2-*ThtAP3-2a* construct consisted of a 427 bp fragment comprising the C-terminal region (227 bp)

and 3'UTR (200 bp) of *ThtAP3-2a*. The TRV2-*ThtAP3-2b* construct consisted of a 408 bp fragment comprising the C-terminal region (253 bp) and 3'UTR (155 bp) of *ThtAP3-2b*. Additional experiments targeting shorter regions (216–342 bp) were attempted first, but they did not trigger sufficient down-regulation. Experiments using double and triple constructs were also attempted unsuccessfully.

Twenty *T. thalictroides* tubers that had been kept in soil in the dark at 4°C for 8 weeks, were treated with each construct as described previously (Di Stilio et al., 2010; Galimba et al., 2012). Briefly, a small incision was cut in the tubers near the bud using a clean razor blade, they were then submerged in infiltration medium containing appropriate *Agrobacterium* cultures (TRV1 and one of the TRV2 constructs), and infiltrated in a chamber under full vacuum (−100 kPa) for 10 min. Nine TRV2-*ThtAP3-1*, five TRV2-*ThtAP3-2a*, and 12 TRV2-*ThtAP3-2b* treated plants survived to flowering. Twenty plants were infiltrated with empty TRV2 vector (mock-treated control, to detect background viral effects). All plants were transferred to the greenhouse following infiltration and were grown together with 20 untreated plants under equal conditions. Leaves and flowers arose approximately 2 weeks later. Flowers from all treatments and controls (untreated and treated with empty vector) were observed, as they opened, under a Nikon SMZ800 dissecting scope, photographed using a Q-Imaging MicroPublisher 3.3 digital camera or a Canon PowerShot SD890 IS digital camera, and flash-frozen for RNA processing.

Molecular Validation of VIGS Lines

After recording their phenotype and photographing, young open flowers that had been flash frozen (as described above) were processed to determine the presence of TRV1 and TRV2 constructs, and the expression levels of B-class genes, including the three *Thalictrum* AP3 genes and the single copy gene *ThtPI*. Total RNA was extracted using TRIzol (Invitrogen, Life Technologies, CA, United States), following manufacturer instructions and DNased using amplification grade DNase (Invitrogen, Life Technologies, CA, United States). First strand cDNA synthesis was carried out on 1 µg of total mRNA using iScript (Bio-Rad, CA, United States), following manufacturer instructions, and diluted 2.5-fold for use in quantitative reverse transcriptase (RT) PCR (qPCR). PCR was carried out with GoTaq (Promega, WI, United States) on 1 µl of cDNA using TRV1- and TRV2-specific primers (Supplementary Table S1) for 30 cycles at 51°C annealing temperature. Locus-specific primers, designed to amplify regions distinct from those used in the virus-induced gene silencing (VIGS) constructs, were used to test expression levels of *ThtAP3-1*, *ThtAP3-2a*, *ThtAP3-2b*, and *ThtPI* by qPCR (Supplementary Table S1). Primers were validated prior to use, using a cDNA dilution series to verify consistent efficiencies and a dissociation curve analysis to confirm single product amplification, as previously published (Galimba et al., 2012). Each qPCR reaction contained 15 µl iQ SYBR Green Supermix (Bio-Rad, CA, United States), 12.2 µl H₂O, 0.9 µl each of forward and reverse primer and 1 µl of cDNA. Samples were amplified on an MJ Research Chromo4 Detector for 45 cycles, in triplicate, including a no-template control. Cycling conditions

¹<https://sites.google.com/a/ualberta.ca/onekp>

were: 94°C for 10 min and 45 cycles of 94°C for 30 s, 53°C for 30 s, and 72°C for 30 s. Relative expression was calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001), and normalized against the averaged expression of two housekeeping genes, *ThtACTIN* and *ThtEF1 α* (*Elongation Factor 1 α*). Since B-class gene expression levels between untreated ($n = 6$) and empty TRV2-treated flowers ($n = 3$) were not significantly different in a Student's two-tailed t -test with unequal variance ($p > 0.05$), both controls were combined for further analyses. The statistical significance of differences in gene expression between controls and treatments was calculated using Student's two-tailed t -test with unequal variance: controls ($n = 9$), TRV2-*ThtAP3-1* ($n = 15$), TRV2-*ThtAP3-2a* ($n = 10$), and TRV2-*ThtAP3-2b* ($n = 4$), and p -values were evaluated using Holm-Bonferroni corrections to avoid Type I error.

Yeast Two-Hybrid Assays

To prepare constructs for yeast two-hybrid analysis, complete coding sequences for *ThtAP3-1*, *ThtAP3-2a*, and *ThtAP3-2b* were cloned into pGADT7 and pGBKT7 vectors using the In-Fusion HD Cloning Kit (Clontech, CA, United States) and custom primers (Supplementary Table S1). The pGADT7 and pGBKT7 constructs for the E-class interacting partner *ThtSEP3 Δ C* were available from a previous study (Galimba et al., 2012). *ThtSEP3 Δ C* is a C-terminal-truncated version of *ThtSEP3* lacking the last 195 bp of coding sequence to avoid auto-activation (Causier and Davies, 2002; Galimba et al., 2012); C-terminal truncation is supported by a number of studies showing that the K domain is more critical for MADS-box protein interactions (Davies et al., 1996; Pelaz et al., 2001; Immink et al., 2009) and is involved both in dimer and tetramer formation (Puranik et al., 2014).

Yeast two-hybrid assays were carried out using Matchmaker Gold Yeast Two-Hybrid System (Clontech, CA, United States). Cells were co-transformed and plated on Leu/Trp-free media to select for diploid colonies. Single colonies were serially diluted 10-fold to 1:10,000 in water and 5 μ l of each dilution were plated on Leu/Trp/His-free media and Leu/Trp/His-free media supplemented with Aureobasidin A (AbA) to test for protein interaction, and on Leu/Trp-free media to control for yeast growth. Interactions were scored and photographed using a Canon PowerShot SD890 IS digital camera after 4 days of incubation at 28°C.

Identification of Promoters and APETALA3 Binding Motif Analysis

In *A. thaliana*, the region extending 496 bp upstream of the *AP3* start codon contains three CArG boxes and drives GUS expression in the same temporal and spatial patterns as the *AP3* transcript in wild-type flowers (Tilly et al., 1998). Using this information as guideline, and to identify promoters in *Thalictrum AP3*, a 1 kb fragment upstream of the start codon of each gene was obtained from a *T. thalictroides* draft genome (unpublished). The *Arabidopsis AP3* frequency matrix was downloaded from JASPAR² (Mathelier et al.,

2014; Figure 5A). This matrix, derived from ChIP-seq data (Wuest et al., 2012), consists of nucleotide frequencies at 15 positions: a 10-nucleotide CArG box surrounded by two 5' and three 3' nucleotides. The use of frequency matrices provides position-specific penalties for deviations from the consensus as opposed to simpler models, like consensus sequences, which treat all mismatches equally (Stormo, 2013). We converted the matrix to a position-specific scoring matrix (PSSM) using the RSAT-convert-matrix tool available at Regulatory Sequence Analysis Tools³ (Medina-Rivera et al., 2015). The *A. thaliana* background model estimation method was utilized with default settings. The PSSM was used in MORPHEUS⁴ (Minguet et al., 2015) to search sections of sequence upstream of the START codon of *ThtAP3-1*, *ThtAP3-2a*, *ThtAP3-2b*, and *ThtPI* for AP3 binding sites. MORPHEUS provides a score for each binding site that is based on the relative affinity of the binding site to the provided matrix, the threshold was set to 5 based on the score distribution histogram (Supplementary Figure S2) to limit the results to the best matching sites.

To visualize the degree of conservation of the *T. thalictroides* putative AP3 binding motifs (found by the method described above), we generated two alignments of the first 500 bp upstream of the start codon of *paleoAP3-1* and *paleoAP3-2* orthologs of other available Ranunculaceae. Sequences were retrieved from GenBank and Phytozome (Supplementary Table S2 and Supplementary Figure S3), aligned using MUSCLE (Edgar, 2004) and CLC Main Workbench 7 (Qiagen), and homologous regions to the *T. thalictroides* AP3 binding sites were identified.

Analysis of Protein Structure

In order to identify the alpha helices known to be critical for MADS box gene functionality (Puranik et al., 2014), and to test whether any of these were lacking in the paralog coding for a truncated AP3 (*ThtAP3-2a*), we analyzed the protein structure of the region comprising the end of the I domain through the C domain of the three *Thalictrum AP3* proteins using Protein Homology/analogy Recognition Engine v 2.0 (Phyre²; Kelley et al., 2015).

RESULTS

Targeted Gene Silencing of Three *T. thalictroides AP3* Orthologs

To investigate the function of three B-class gene orthologs of the ranunculid *T. thalictroides*, we conducted targeted VIGS using single gene constructs, analyzed gene expression of treated plants in relation to untreated and empty vector controls, and compared the resulting floral phenotypes.

Flowers treated with empty TRV2 vectors and verified for the presence of TRV1/2 transcripts (Supplementary Figure S1) were combined with untreated wild-type flowers, and used as a control treatment for molecular validation ($n = 9$

²<http://jaspar.genereg.net>

³<http://www.rsat.eu>

⁴<http://biodev.cea.fr/morpheus/>

flowers from four plants; **Figure 2**). Wild-type *T. thalictroides* flowers are apetalous, with 5–12 white or pink petaloid sepals enclosing 45–76 filamentous stamens and 3–11 free carpels with prominent papillae at the stigma (Di Stilio et al., 2009; Galimba et al., 2012; **Figure 3A**). Wild-type leaves are compound with lobed leaflets (**Figure 3B**). Empty TRV2-treated flowers showed comparable morphology to wild-type (**Figure 3C**). Small sections of brown necrotic tissue were present on 19% of treated flowers, similar to previously reported background viral effect (Di Stilio et al., 2010), and these were therefore not counted as phenotype.

Flowers treated with TRV2-*ThtAP3-1* and verified for the presence of TRV1/2 transcripts ($n = 15$ flowers from four plants; Supplementary Figure S1) exhibited a significant down-regulation of *ThtAP3-1* transcripts, with an average 5.3-fold decrease in gene expression levels with respect to controls ($p < 0.001$, **Figure 2A**). Down-regulation of *ThtAP3-1* at the individual level ranged from a 1.5- to 32-fold decrease in gene expression (**Figure 2A**). Average expression levels of the other three B-class genes were statistically like controls (**Figures 2B–D** and Supplementary Table S3). All flowers ($n = 24$) from validated plants were phenotyped (**Figure 3S**). Abnormal phenotypes included narrow sepals in 38% of flowers (**Figure 3D**), smaller and possibly extra inner sepals, against the outer stamens in 38% of flowers (**Figure 3E**, sepal number is variable in *Thalictrum*, but positioning and size of these is distinct) and chimeric organs (sepal/stamen or sepal/carpel) in the sepal/stamen boundary in 13% of flowers (**Figures 3E–H**). These chimeric organs included sepaloid organs with anther sacs (**Figures 3E,F**) or sepaloid organs with stigma-like tissue on the distal end (**Figures 3G,H**). We also observed stunted stamens with small anthers in 42% of flowers and/or short filaments in 54% of flowers (**Figures 3D,G,I**). Stunted stamens were not present in our empty TRV2-treated flowers, although we have observed them previously at low frequencies in viral controls (Soza et al., 2016). TRV2-*ThtAP3-1* treated plants also possessed small sepals on 50% of flowers and/or curved sepals on 63% of flowers (**Figures 3G,I**). We also observed patches of green tissue on sepals of 13% of flowers (**Figure 3I**) and smooth stigmas, due to the absence of stigmatic papillae, on 29% of flowers (**Figures 3I,J**). In summary, in addition to the expected role in stamen identity, down-regulation of *ThtAP3-1* altered sepal size (and possibly number), shape and color, changed organ identity at the sepal/stamen boundary and resulted in loss of stigmatic papillae.

Treatment with TRV2-*ThtAP3-2a* failed to down-regulate *ThtAP3-2a* in flowers verified for the presence of TRV1 and TRV2 transcripts ($n = 10$ flowers from three plants; Supplementary Figure S1). On the contrary, it resulted in up-regulation of all *ThtAP3s* and *ThtPI* (**Figures 2A–D**). These flowers showed, on average, a 2.9-fold increase in *ThtAP3-2a* expression, which was significantly higher than controls ($p < 0.001$; **Figure 2B**). Up-regulation of *ThtAP3-2a* at the individual level ranged from a 1.2- to 4.7-fold increase in gene expression (**Figure 2B**). They also had, on average, significantly higher gene expression levels than controls for *ThtAP3-1* (2.1-fold, $p = 0.005$; **Figure 2A** and Supplementary Table S3) and *ThtPI* (1.8-fold, $p = 0.013$;

Figure 2D and Supplementary Table S3). *ThtAP3-2b* expression levels were 1.8-fold higher in TRV2-*ThtAP3-2a* treated flowers; this increase was not significant when applying Holm–Bonferroni corrections ($p = 0.026$; **Figure 2C** and Supplementary Table S3). Flower ($n = 15$) phenotypes from validated plants consisted of a subset of those observed in *ThtAP3-1* treated plants (**Figure 3S**). Sepal/stamen chimeric organs in place of outer stamens were present on 7% of flowers (**Figure 3K**), and small extra inner sepals were present on 20% (**Figure 3L**). Lobed sepals were present on 33% of treated flowers (**Figure 3M**) and were not present in any of the untreated plants. We observed curved sepals on 33% of flowers and stunted stamens with short filaments and/or small anthers on 13% of flowers. We did not observe narrow, small, or green outer sepals in this treatment. Small, underdeveloped carpels with smooth stigmas, including one with an exposed ovule (**Figure 3N**), were present in 13% of flowers. In summary, we were unable to detect down-regulation of *ThtAP3-2a* after treatment with TRV2-*ThtAP3-2a*, and instead detected unexpected up-regulation of all B-class genes (albeit not statistically significant for *ThtAP3-2b*), which was associated with phenotypes affecting stamen/sepal boundary, sepal morphology, and carpel development.

Flowers treated with TRV2-*ThtAP3-2b* and verified for the presence of viral transcripts ($n = 4$ flowers from three plants; Supplementary Figure S1) showed, on average, a strongly significant down-regulation of *ThtAP3-2b* compared to controls (2.8-fold decreased expression, $p = 0.001$; **Figure 2C**). *ThtAP3-2b* was down-regulated a 4.1- to 7.4-fold at the individual level (**Figure 2C**). All other B-class gene expression levels were like controls (**Figures 2A,B,D** and Supplementary Table S3). Flowers ($n = 27$) from validated plants (**Figure 3S**) showed lobed sepals on 26% of flowers, with green sectors on 7% of flowers (**Figure 3O**). Small sepals were also present on 48% of flowers and curved sepals were present on 30%. We also observed stamens replaced with chimeric organs (sepaloid with carpeloid or stamenoid features) on 15% (**Figures 3P,Q**) and small extra inner sepals on 15% of flowers (**Figure 3P**). Plants also displayed narrow sepals on 19% of flowers (**Figure 3R**) and stunted stamens, with short filaments on 30% of flowers and/or small anthers on 11% (**Figure 3R**). Smooth stigmas with reduced papillae were present on 4% of flowers. In summary, down-regulation of *ThtAP3-2b* altered sepal size, shape and color, stamen identity, sepal/stamen boundary, and stigmatic papillae development in a similar manner as the down-regulation of *ThtAP3-1*, with the addition of sepal lobing.

Taken together, phenotypes resulting from VIGS treatments (and not observed in control flowers) affected either sepal, stamen, or carpel morphology. Abnormal sepal morphology occurred in flowers across treatments: down-regulation of *ThtAP3-1* and *ThtAP3-2b* resulted in reduction in overall size and width (small and narrow), shape (curved instead of flat) and color (green), with *ThtAP3-2b* VIGS flowers also exhibiting sepal lobing. Similar lobing and curving of sepals, was observed in TRV2-*ThtAP3-2a* treated flowers, which exhibited higher expression of all B-class genes, although none of the additional sepal defects were present (narrower, smaller, or green). The second major category of phenotypes

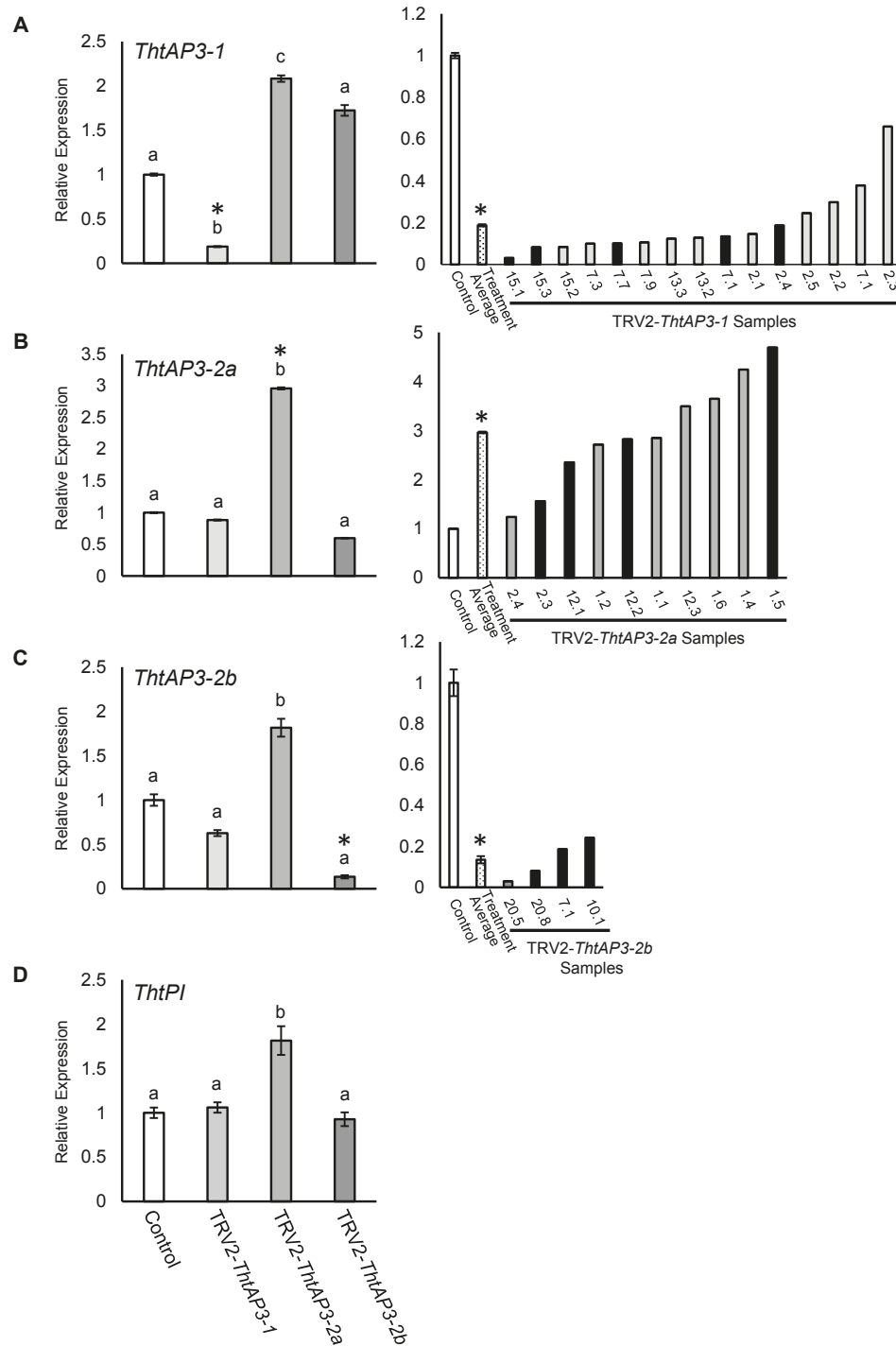


FIGURE 2 | Molecular validation of targeted gene silencing of *Thalictrum thalictroides* AP3 orthologs. **(A–D)** Left panels: Quantitative RT-PCR (qPCR) showing average relative expression levels of *ThtAP3-1* **(A)**, *ThtAP3-2a* **(B)**, *ThtAP3-2b* **(C)**, and *ThtPI* **(D)** in each of three treatments, TRV2-*ThtAP3-1*, TRV2-*ThtAP3-2a*, and TRV2-*ThtAP3-2b* by virus-induced gene silencing. Control ($n = 9$: untreated, $n = 6$ plus TRV2-empty, $n = 3$); *ThtAP3-1* VIGS ($n = 15$); *ThtAP3-2a* VIGS ($n = 10$); and *ThtAP3-2b* VIGS ($n = 4$). Gene expression levels of the control were set to 1. Expression was normalized against *ThtACTIN* and *ThtEEF1*. Mean and SE of biological replicates are shown; different letters indicate a significant difference in Student's two-tailed t -test with unequal variance and Holm–Bonferroni corrections ($p \leq 0.05$). **(A–C)** Right panels: Molecular validation of targeted gene silencing in individual flowers (labels represent plant number, followed by flower number) relative to control (set to 1). Treatment average (dotted bar) of individual flowers (biological replicates). SE shown for control and treatment average values. Black bars indicate flowers shown in **Figure 3**. Asterisks cross-reference same treatment averages compared between treatments (left), or amongst individual samples within a treatment (right).

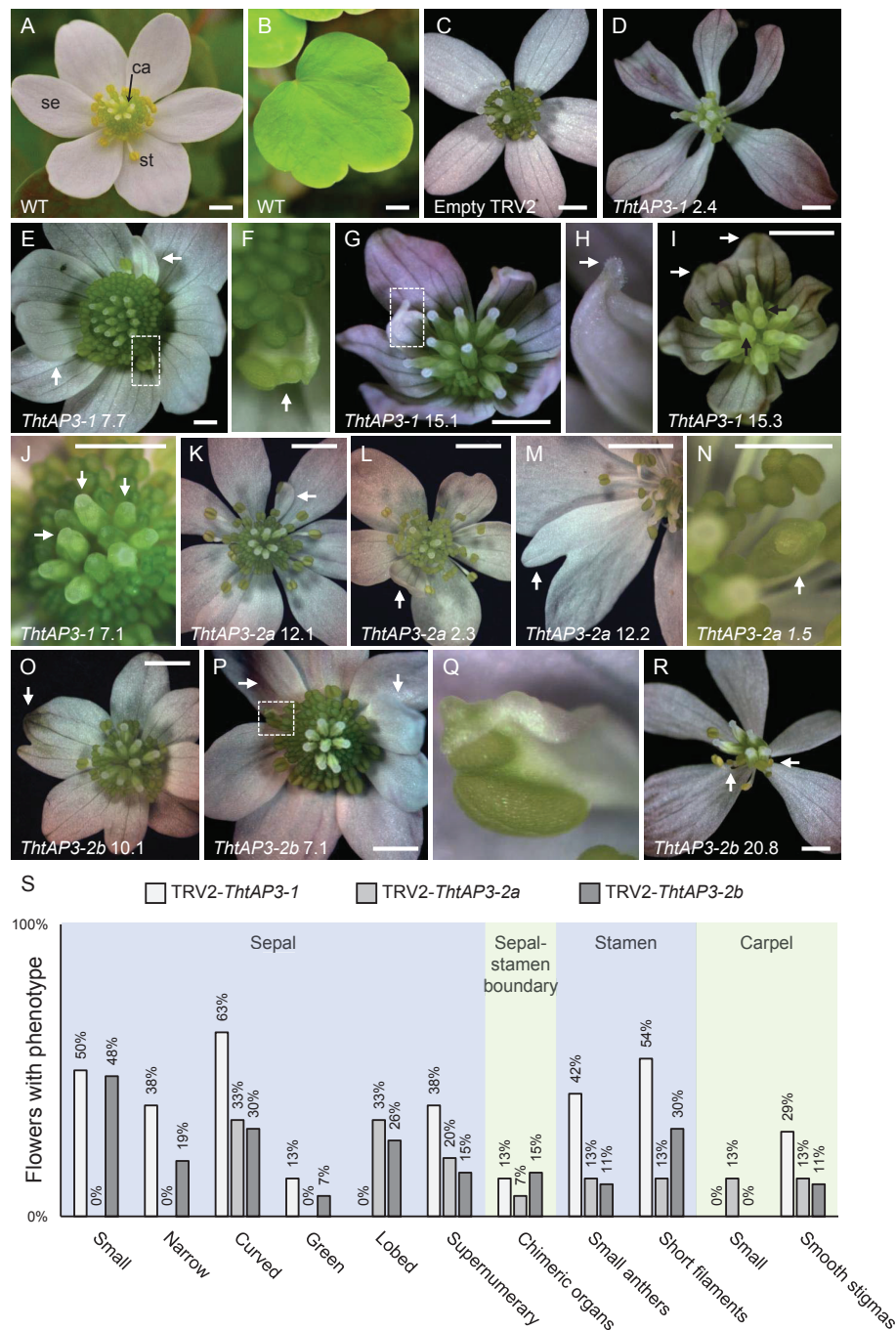


FIGURE 3 | Phenotypes resulting from targeted silencing of *Thalictrum thalictroides* AP3 orthologs. **(A)** Untreated *T. thalictroides* representative flower. **(B)** Untreated representative leaflet. **(C)** Flower treated and validated for empty TRV2 vector. **(D–J)** Representative flowers treated with and validated for TRV2-*ThtAP3-1*. **(D)** Flower with narrow sepals. **(E)** Flower with chimeric organ (boxed) and extra inner sepals (arrows). **(F)** Detail of sepal/stamen chimera in panel **(E)** (arrow). **(G)** Flower with small, curved sepals and chimeric organ in place of outer stamen (boxed). **(H)** Detail of sepal–carpel chimera in panel **(G)**, with stigma (arrow). **(I)** Flower with small, curved, greener sepals, green segments indicated by white arrows, and smooth stigmas, indicated with black arrows. **(J)** Flower with smooth stigmas, three indicated with arrows. **(K–N)** Representative flowers treated with TRV2-*ThtAP3-2a*. **(K)** Flower with chimeric sepal/stamen organ in place of outer stamen (arrow). **(L)** Flower with extra inner sepal (arrow). **(M)** Flower with lobed sepal (arrow). **(N)** Flower with unfused carpel and exposed ovule (arrow). **(O–R)** Representative flowers treated with TRV2-*ThtAP3-2b*. **(O)** Flower with lobed sepal with green sectors (arrow). **(P)** Flower with chimeric organ (boxed) and extra inner sepals (arrows). **(Q)** Detail of sepal/stamen chimera in panel **(O)**. **(R)** Flower with narrow sepals and stunted stamens exhibiting small anthers and short filaments (arrows). **(S)** Percentage of flowers exhibiting each phenotype (from plants verified for TRV1/2 transcripts): TRV2-*ThtAP3-1* ($n = 24$); TRV2-*ThtAP3-2a* ($n = 15$); and TRV2-*ThtAP3-2b* ($n = 27$). Sample labels represent plant number, followed by flower number (see **Figure 2** for individual expression levels of target genes). Scale bar = 1 mm.

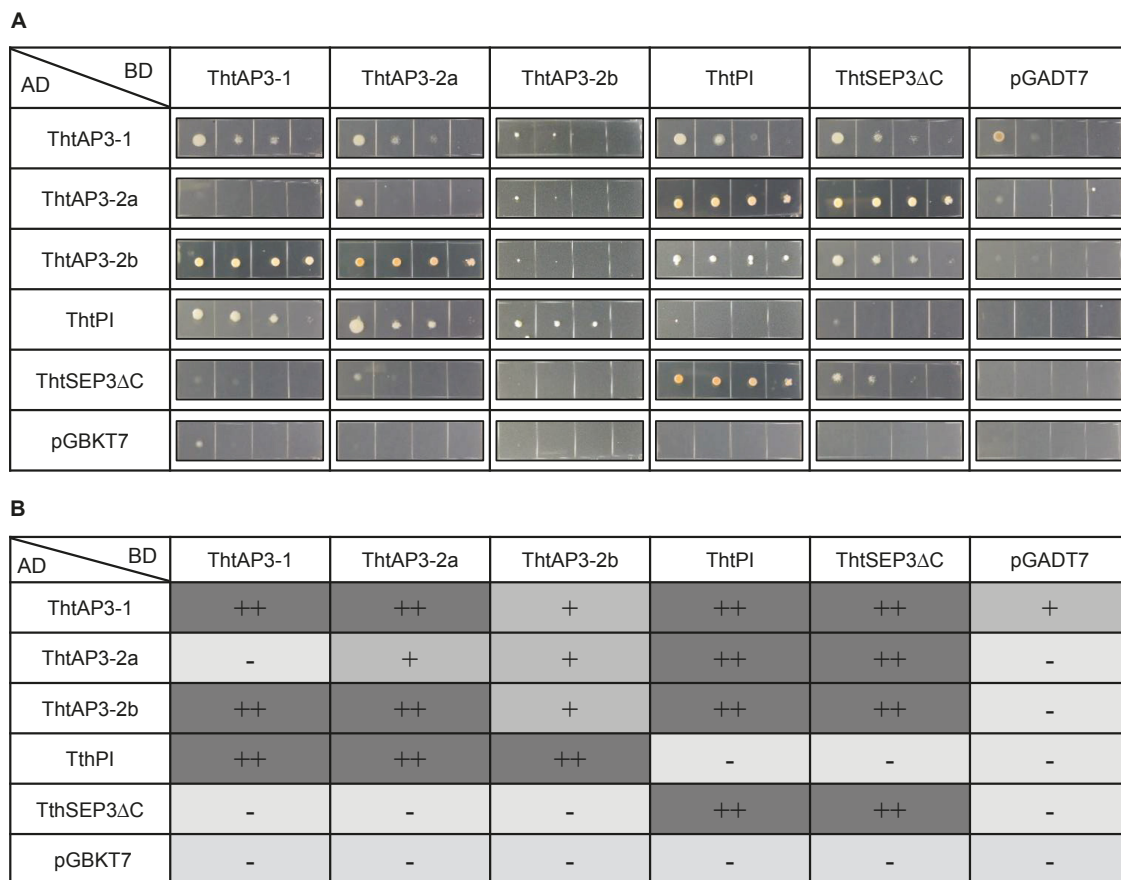


FIGURE 4 | Protein–protein interactions among B-class gene products of the ranunculid *Thalictrum thalictroides*. Interactions between the three ThtAP3 proteins, ThtPI and ThtSEP3 of *T. thalictroides* were determined with the yeast two-hybrid system. The C-terminus of ThtSEP3 was truncated to avoid autoactivation (ThtSEP3ΔC). **(A)** Colony growth on selective Leu/Trp/His-free + AbA medium. Yeast cells were spotted in four 10-fold serial dilutions (from left to right) for each interaction tested. All proteins were expressed as fusion with the GAL4 activation domain (AD) and the GAL4 DNA binding domain (BD). pGBKT7 and pGADT7 are empty vector controls. **(B)** Interpretation of the strength of the protein–protein interactions shown in panel **(A)**. –, no interaction; +, interaction; ++, strong interaction.

affected stamen morphology across the three treatments: stamens looked stunted, had small anthers, their filaments failed to properly elongate, and outer stamens were sometimes replaced by chimeric organs or small sepals. Lastly, down-regulation of *ThtAP3-1* and *ThtAP3-2b* affected late carpel development, leading to the absence of stigmatic papillae, while the high expression of B-class genes in TRV2-*ThtAP3-2a* treated flowers caused carpel stunting and failure to fuse around the ovule.

Yeast Two-Hybrid Analyses: Promiscuous Interactions Among *Thalictrum* B-Class Proteins

To test the interactions among the four B-class proteins present in *T. thalictroides*, we performed yeast two-hybrid analyses, including the predicted E-class partner protein ThtSEP3 (Figure 4). We used a ThtSEP3ΔC construct with a truncated C-terminus to avoid previously documented auto-activation (Galimba et al., 2012). Control transformations with empty

vectors (pGBKT7 and pGADT7) produced minimal to no growth, ruling out auto-activation for the other proteins.

As expected, all three ThtAP3 proteins could dimerize with their B-class partner ThtPI, and this interaction was observed in both directions. A more unexpected result was that all three ThtAP3 proteins could heterodimerize with each other: ThtAP3-1 interacted with ThtAP3-2a and ThtAP3-2b, and ThtAP3-2a interacted with ThtAP3-2b. These interactions varied in strength, and ThtAP3-1 and ThtAP3-2a had a positive interaction in one direction and a negative one in the other. ThtAP3-1 interacted weakly with the empty vector in one direction, and this may decrease support for the weak interaction detected between ThtAP3-1 and ThtAP3-2b. Using the yeast culture dilutions (Figure 4A) to interpret the strength of the protein–protein interactions (Figure 4B), we observed that ThtAP3-1 could homodimerize strongly, while ThtAP3-2a and ThtAP3-2b homodimerized weakly and ThtPI did not form homodimers. All four B-class proteins (ThtAP3-1, ThtAP3-2a, ThtAP3-2b, and ThtPI) could interact with the E-class partner ThtSEP3. Taken together, all three ThtAP3

proteins heterodimerize with ThtPI and ThtSEP3, as expected, and additionally homodimerize and heterodimerize with each other.

Identification of Putative AP3 Binding Motifs in *T. thalictroides* B-Class Gene Promoters

To find additional evidence for auto- and/or cross-regulation of B-class genes by ThtAP3 proteins, we searched for AP3 binding sites in *Thalictrum* B-class gene promoter regions using a frequency matrix (Figure 5A) derived from experimental binding assays that includes AP3-specific binding frequencies to the CArG box plus five additional nucleotides (Sharma et al., 2011; Sharma and Kramer, 2013), as implemented in MORPHEUS (Minguet et al., 2015). All binding site sequences identified above the threshold (>5 , Supplementary Figure S2) are listed in Supplementary Table S4. Three putative AP3 binding sites were identified for *ThtAP3-1*, located at positions (upstream from start codon) -262 , -139 , and -121 (Figure 5B). *ThtAP3-2a* had only one binding site (position -9 ; Figure 5C), *ThtAP3-2b* had three (positions -293 , -151 , and -78 ; Figure 5D), and *ThtPI* had one (position -340 ; Figure 5E). Binding sites at position -121 in the *ThtAP3-1* promoter, -91 in the *ThtAP3-2a* promoter, and -78 in the *ThtAP3-2b* promoter appear to be homologous based on our alignments with additional ranunculids (Supplementary Figure S3), and are the most conserved. They are 86.7% identical between *ThtAP3-1* and *ThtAP3-2b*, 66.7% between *ThtAP3-1* and *ThtAP3-2a*, and 80% between *ThtAP3-2a* and *ThtAP3-2b* (purple boxes in Figure 5F). The other binding sites are locus-specific (Figure 5F). The alignment to orthologs from other Ranunculaceae also shows the conservation of these binding sites within the family, and the sequence divergence between *ThAP3-2a* and *RanAP32-b* (Supplementary Figure S3). The AP3-2 alignment reveals that two of *ThtAP3-2b* motifs have corresponding homologous sequence in *ThtAP3-2a* (Supplementary Figure S3B), yet they have diverged enough that their MORPHEUS scores (-1 and -3.1) were below the set threshold (Figure 5F, shown in lighter shading), suggesting that they are no longer functional.

Protein structural analysis on amino-acid translations of the coding sequences of all three AP3 loci identified the two alpha helices spanning the I and K domains (Supplementary Figure S4). These helices have been shown to be critical for SEP3 multimerization, and possibly for MADS box protein interactions more generally (Puranik et al., 2014). Despite its C-terminus truncation, *ThtAP3-2a* appears to possess the required helices for dimer and tetramer formation. The first helix, involved in dimer formation, has six amino acid disparities between the two paralogs; the second helix, involved in tetramer formation (Puranik et al., 2014), has 12 amino acid disparities. Of these, five involve hydrophobic residues in one paralog or the other, and a broader alignment across multiple ranunculid taxa revealed that three of these are specific to AP3-2a protein orthologs (Supplementary Figure S4). Further functional testing, combined with protein crystallography, would be necessary to ascertain potential effects of divergent residues in mediating dimerization

or tetramerization between *ThtAP3-2a* and other MADS-box proteins.

DISCUSSION

This study investigated the function of three AP3 orthologs representing ancient as well as recent gene duplications in *T. thalictroides*, a representative of the sister lineage to core eudicots. We enquired whether these gene paralogs have remained redundant or diverged in function, and whether they contribute to ectopic petaloidy of sepals. Down-regulation of two of the paralogs (*ThtAP3-1* and *ThtAP3-2b*) correlated with sepal, stamen, and stigma developmental defects. In addition to the expected interaction with the E-class protein ThtSEP3 and the other member of the B-class lineage ThtPI, all ThtAP3s had the ability to form homodimers and to interact with each other, providing a potential mechanism for novel function. Putative binding motifs were identified in all B-class gene promoters, supporting the potential for cross-regulatory interactions. Based on the evidence presented here, we conclude that, in addition to a conserved role in stamen identity, *ThtAP3-1* and *ThtAP3-2b* contribute to ectopic petaloidy of sepals in this species, as evidenced by their effect on sepal color, shape, and size. They also appear to regulate the genetic pathway leading to stigmatic papillae in carpels. Finally, we propose two working hypotheses to explain the unexpected up-regulation of all B-class genes in plants targeted for silencing of *ThtAP3-2a*, which codes for a truncated protein.

Prior characterization of *Thalictrum* AP3 expression found high expression of all three loci in sepals and stamens of *T. thalictroides*, yet little to no expression in the reduced green sepals of *T. dioicum* (Di Stilio et al., 2005; LaRue et al., 2013). Together with the phenotypes emerging from our VIGS experiments, these data suggest a combined function of *ThtAP3* genes in stamen identity and in the partial transference of petaloid features to perianth organs, otherwise described as sepals based on morphology and evolutionary context (e.g., they have multiple vascular strands, they completely enclose the floral bud throughout development, and petaloid sepals are present throughout the Ranunculaceae). Even though there is no prior expression data for stigmas, our VIGS data suggest that *ThAP3* influence the development of stigmatic papillae. The conical/papillate-cell identity gene *MIXTA* is under the control of B-class genes in *Antirrhinum* and *Phalaenopsis* (Perez-Rodriguez et al., 2005; Manchado-Rojo et al., 2012; Pan et al., 2014) and is expressed in stigmatic papillae of *Thalictrum* (Di Stilio et al., 2009). The loss of stigmatic papillae in carpels of flowers experiencing down-regulation of *ThtAP3-1* and *ThtAP3-2b* in our VIGS experiments suggests that B-class regulation of *MIXTA* possibly also occurs in *Thalictrum*.

Petaloid appearance can have different genetic underpinnings, with certain species relying on B-class gene expression for ectopic petaloidy (e.g., tulips; van Tunen et al., 1993; Kanno et al., 2003), while others exhibit petaloid morphology independent of B-class gene expression (e.g., *Aristolochia*; Pabón-Mora et al.,

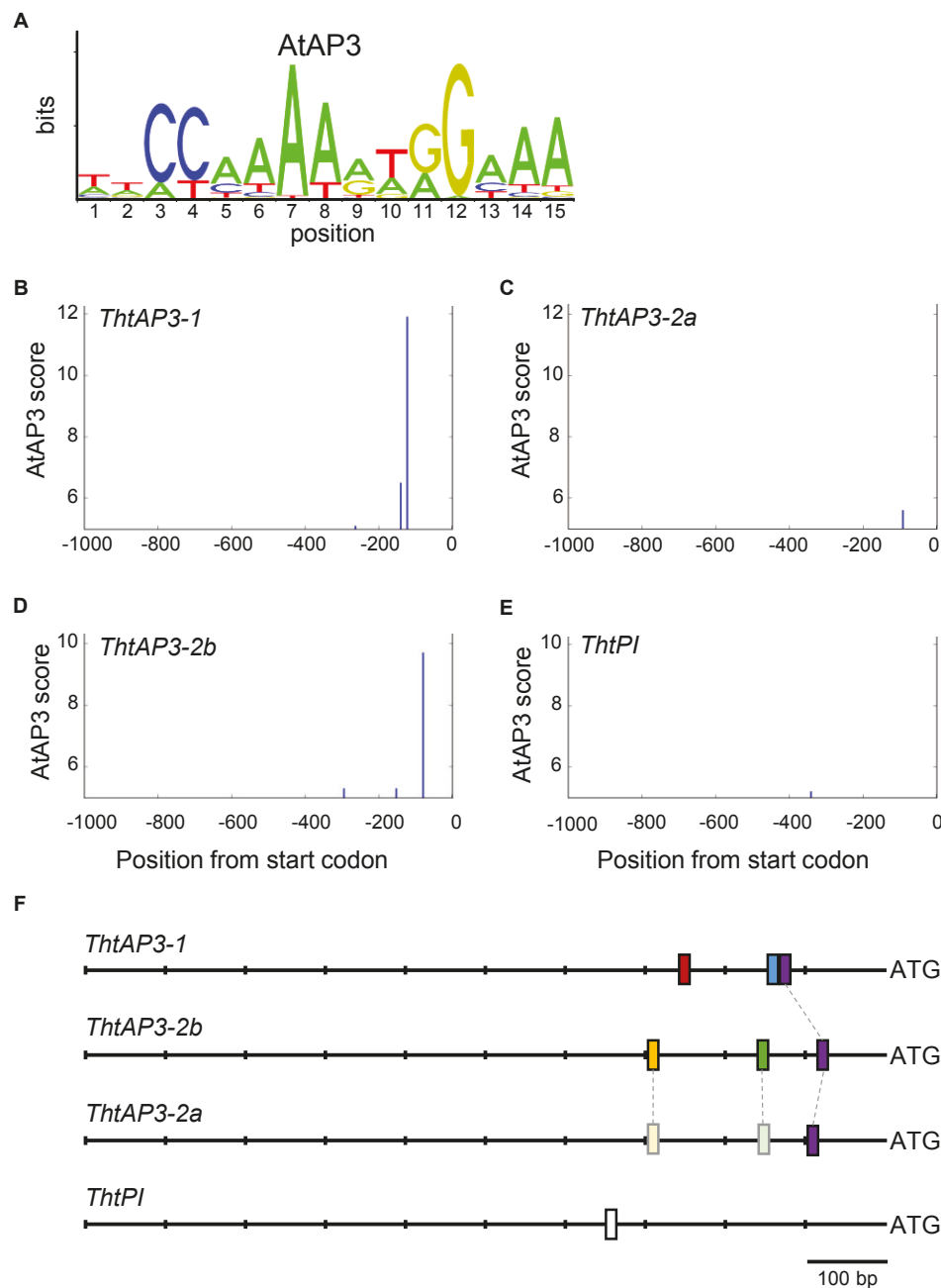


FIGURE 5 | Location of APETALA3-like binding sites on *Thalicttrum thalictroides* B-class gene promoters. **(A)** Sequence logo for *Arabidopsis* AP3 binding frequency matrix (jaspar.genereg.net), with letter sizes relative to nucleotide frequency. **(B–E)** Positions identified by MORPHEUS as AP3 binding sites, in 1 kb sections of sequence directly upstream of each gene. The height of each vertical bar indicates the MORPHEUS score of that binding site, with the cutoff score set at 5. **(F)** Schematics of promoter regions for each gene, with binding site positions indicated by bars. A putative homologous binding site shared by all three *ThtAP3* promoters is shown in the same color (purple), and two putative homologous regions in *ThtAP3-2a* are shown in lighter shades (see also Supplementary Figure S3).

2015). E-class genes are also known to contribute to petaloid features of sepals in *Thalicttrum* (Soza et al., 2016). Given the presence of petaloid sepals throughout Ranunculaceae, it is unlikely that they arose *de novo* in *Thalicttrum*. However, *Thalicttrum* AP3s have a distinct role in petaloid morphology

of sepals when compared to other ranunculids where it has been characterized: *Aquilegia* (Sharma and Kramer, 2013) and *Nigella* (Wang et al., 2015) in the Ranunculaceae and *Papaver* (Drea et al., 2007) in the Papaveraceae. For example, in *Aquilegia*, loss of *AqPI* function results in loss of anthocyanin

production in sepals, but does not alter other aspects of sepal morphology (Sharma and Kramer, 2017). Green streaks on sepals in our VIGS experiments are comparable to those found in sepals of *Nigella damascena* loss-of-B-function mutants, which also lost anthocyanin pigmentation. In spite of this, B genes were not interpreted as playing a role in sepal petaloidy in that system (Wang et al., 2015). In *Thalictrum*, we interpreted a change toward a more leaf-like appearance of perianth organs, such as change in shape, reduction in size and the gain of photosynthetic pigment, as loss of “petaloid” features. The fact that AP3 and PI negatively regulate genes involved in chlorophyll accumulations resulting in the white *A. thaliana* petals (Mara et al., 2010), is consistent with our observations of sepal greening during down-regulation of B-class genes in *Thalictrum*, and points to similar mechanisms for this petaloid feature in petals and sepals. Based on evidence presented here, we propose that the common ancestor of *Thalictrum* evolved novel interactions among B-class proteins following its lineage-specific *ThtAP3-2a/b* gene duplication. This, in turn, enabled new regulatory interactions that contributed to sepal petaloidy in conjunction with the E-class partners (Soza et al., 2016). Consistent with this hypothesis, the wind-pollinated species *T. dioicum* has small, green sepals that do not express the *AP3-2* loci early in development (Di Stilio et al., 2005).

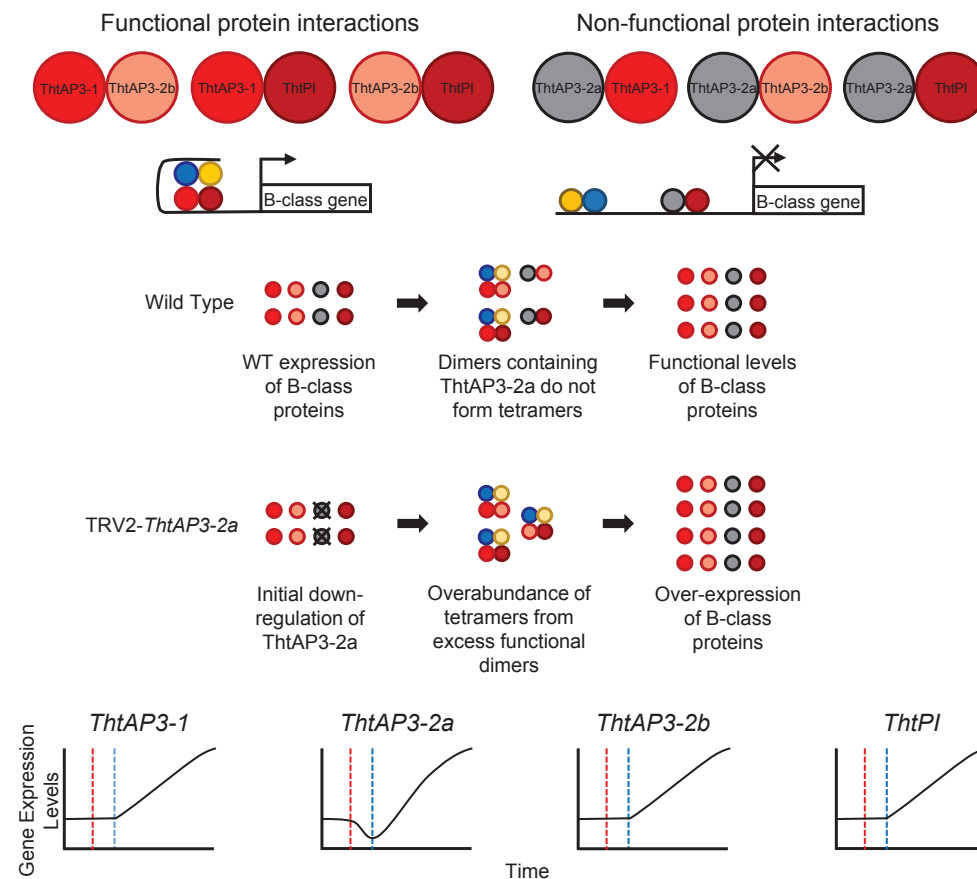
Given that orthologs of the A-class gene *APETALA1* do not exist in the order Ranunculales (Litt and Irish, 2003), it is likely that the genetic factors involved in *Thalictrum* sepal identity differ from those in the evolutionarily derived core eudicots. For example, the presence of chimeric outer stamens with combined carpel and sepal features in TRV2-*ThtAP3-1* and TRV2-*ThtAP3-2b* treated flowers could be considered inconsistent with the mutually antagonistic nature of the A- (sepal identity) and C- (carpel identity) classes (Bowman et al., 1991), as defined in the original ABC model. Alternatively, the presence of ectopic carpeloid features may suggest that, like in *Arabidopsis*, *T. thalictroides* B-class genes act as suppressors of carpel development genes outside of the carpel zone (Wuest et al., 2012). In *Nigella*, AP3 homologs function in keeping the stamen–petal boundary (Gonçalves et al., 2013; Wang et al., 2015). The presence of sepal/stamen intermediates in the boundary region between these two organs in *Thalictrum* knockdowns suggests that, in the absence of petals and of the *AP3-3* petal identity gene, the other AP3 paralogs perform a boundary-keeping role between the stamen and perianth zones (sepals, in this case). This observation is consistent with a role of B-class genes in regulating the expression zone of the C-class gene *AG* in other ranunculids (Lange et al., 2013; Sharma and Kramer, 2017). No carpel–sepal chimeras were observed in plants targeted for *ThtAP3-2a* silencing, which were actually over-expressing B-class genes. TRV2-*ThtAP3-2a* treated flowers also never exhibited signs of loss of petaloidy, such as smaller, narrower, or green sepals, and they were the only treatment to cause stunted and unfused carpels, a potential sign of B function disrupting normal carpel identity in the inner floral zone. Presence of other phenotypes, such as stamen–sepal chimeras and lobed sepals, may be explained by an

overall disruption of the protein ratios necessary for normal development.

Down-regulation of the *ThtAP3* paralogs individually did not result in complete homeotic conversion of stamens into carpels or complete loss of petaloidy, suggesting at least a partial degree of redundancy. This is unlike the single copy PI ortholog *ThtPI*, which results in complete loss of stamen identity and a full conversion to carpels upon down-regulation (LaRue et al., 2013). Loss of petaloid features was also partial in *ThtPI* knockdowns, consisting of green streaks on otherwise white sepals, suggesting partial redundancy of this function with other B-class genes, and likely also E-class genes (LaRue et al., 2013; Soza et al., 2016). Incomplete homeotic conversion of floral organs can be attributed to partial redundancy among the *ThtAP3* paralogs, supported by the presence of similar phenotypes in different treatments. This scenario differs from *Aquilegia*, in which the three *AqAP3* paralogs have sub-functionalized to stamen or petal identity (petals are absent in *Thalictrum* and so is the petal-identity paralog *AP3-3*) and neo-functionalized to staminodia identity (a fifth type of organ not present in *Thalictrum*) (Sharma et al., 2011; Sharma and Kramer, 2013). Additional experiments using double and triple constructs to target multiple genes will be needed to fully dissect the degree of redundancy amongst the *Thalictrum* AP3 paralogs.

The up-regulation of all B-class genes in plants targeted for *ThtAP3-2a* silencing was unexpected and, barring a general issue with our VIGS experiments, we propose two (admittedly speculative) working hypotheses leading to testable predictions for future experiments. The truncated ThtAP3-2a protein could be acting as a dominant negative (**Figure 6A**), or there could be a unidirectional back-up circuit effect (**Figure 6B**). Under the dominant negative hypothesis, ThtAP3-2a is able to form protein dimers (**Figure 4A**), but unable to form tetramers. A dominant negative regulation may arise from mutations in ThtAP3-2a affecting hydrophobic residues in the K helices (Supplementary Figure S4) that are key to tetramerization (Puranik et al., 2014), or from the loss of the highly conserved C-terminal motifs (the paleo-AP3 region and PI motif-derived; Kramer et al., 1998). While domain-swap and rescue experiments in *Arabidopsis*, including a truncated *Chloranthus* AP3, have demonstrated that the C domain and its motifs do not affect protein function (Piwarzyk et al., 2007; Su et al., 2008), the ability for the ranunculid *Eschscholzia californica* PI ortholog SEIRENA to form tetramers is dependent on five conserved residues within the PI motif (Lange et al., 2013). One future direction to test whether ThtAP3-2a is acting as a dominant negative would be to overexpress it, with the prediction that it would result in a loss of B-function phenotype; unfortunately stable transformation protocols are currently not available for *Thalictrum*. Alternatively, back-up circuits have been proposed as a mechanism in paralog retention, where one paralog compensates for the decreased expression of a partner gene (**Figure 4B**; Kafri et al., 2005, 2006). The variable distribution of AP3 binding sites among the three *ThtAP3* promoters (**Figure 5F**) lends support to this hypothesis, as paralogs with partially overlapping regulatory motifs are more efficient at

A Dominant Negative Model



B Back-up Circuit Model

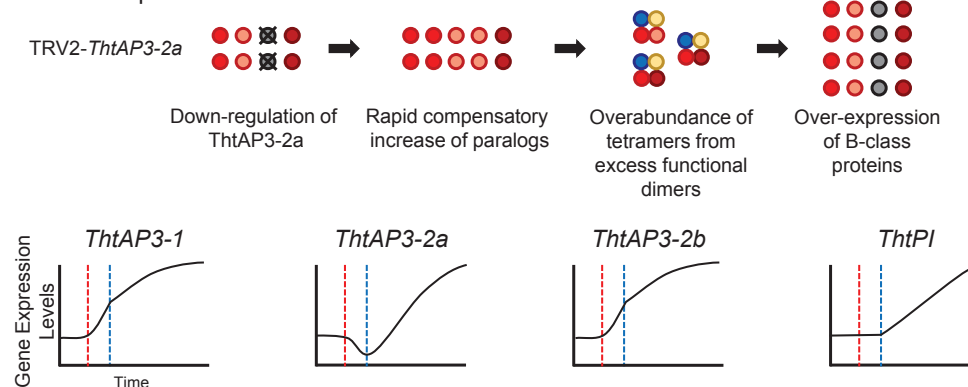


FIGURE 6 | Hypothetical models for B-class gene regulation in *Thalictrum thalictroides*. **(A)** Dominant Negative Model. ThtAP3-2a is presumably able to form protein dimers (K domain intact), either hetero-dimers (shown) or homo-dimers (not shown). These “non-functional” dimers are unable to form the tetramers needed to drive transcription, as required for positive auto- and cross-regulation. ThtAP3-1 and ThtAP3-2b form protein complexes that positively auto- and cross-regulate, driving transcription of the other B-class genes. After *ThtAP3-2a* is down-regulated, as a result of targeted gene silencing (red line in graphs), the higher ratio of functional to non-functional dimers creates an overabundance of tetramers, which drives the over-expression of all B-class genes (blue line). **(B)** Back-up Circuit Model. In response to the initial down-regulation of *ThtAP3-2a* (dotted red line in graphs) during targeted gene silencing, expression of *ThtAP3-1* and *ThtAP3-2b* is up-regulated by a (as yet unknown) compensatory mechanism. Higher expression of *ThtAP3-1* and *ThtAP3-2b* initiates an excess of positive auto and cross-regulation due to total protein levels, as opposed to protein ratios in the Dominant Negative Model (note different shapes of curves from panel **A**), causing an increase in the expression of all B-class genes (dotted blue line).

rescuing the function of their mutated counterpart than paralogs with highly similar or highly dissimilar sets of motifs (Kafri et al., 2005). Likewise, the presence of more AP3 binding sites in the promoters of *ThtAP3-1* and *ThtAP3-2b* supports the hypothesis that these genes are “backing-up” *ThtAP3-2a*. Both hypothetical models rely on an initial *ThtAP3-2a* down-regulation to trigger an increase in the positive regulation of all B-class genes. The back-up model presumably leads to a more immediate increase in expression of *ThtAP3-1* and *ThtAP3-2b*, as there is also a primary “back-up” response of these paralogs (by an unknown mechanism, note different shape of gene expression curves in **Figure 6A** versus **Figure 6B**). Since we were unable to detect the putative transient down-regulation of *ThtAP3-2a*, a more comprehensive expression analysis would need to be done during earlier developmental stages to provide evidence for either of these arguments. Additional experiments are also needed to test identified binding sites *in silico*, and to ascertain whether divergent amino acids or lack of C-terminal motifs negatively affect *ThtAP3-2a* function.

As a representative of an early diverging eudicot, *Thalictrum* is in a key phylogenetic position to study the evolution of floral MADS box protein–protein interactions. Our yeast two-hybrid assays showed novel and promiscuous interactions among the different *Thalictrum* AP3s. Core eudicot EuAP3 and most monocot paleoAP3 proteins do not homodimerize, but rather interact as obligate heterodimers with PI (Schwarz-Sommer et al., 1992; Tröbner et al., 1992; Riechmann et al., 1996; Moon et al., 1999; Vandenbussche et al., 2004), and this may have led to the canalization and increased robustness of the eudicot flower (Lenssen et al., 2009; Melzer et al., 2014). However, AP3 and PI orthologs from early-diverging angiosperms and monocots are able to interact both as homo- and heterodimers (Melzer et al., 2014). AP3 homodimerization was probably lost very early in angiosperm evolution (before the eudicot–monocot split), while PI homodimerization was likely lost later, but before the diversification of the eudicots (Melzer et al., 2014). Evidence presented here suggests that *Thalictrum* floral MADS-box protein biochemical behavior more closely matches that of early diverging angiosperms than that of core eudicots. *Aquilegia vulgaris* AP3 (AqvAP3-2 and AqvAP3-3) have also been shown to homodimerize, yet the three proteins do not dimerize with each other (Kramer et al., 2007). All three *Thalictrum* AP3s not only heterodimerize with ThtPI as expected, but also homodimerize and heterodimerize with each other (**Figure 4**). In addition, all four B-class protein products can interact with the E-class partner ThtSEP3. While certain interactions only occurred in one direction, similar asymmetrical results have been observed, and deemed valid, in previous publications for other MADS box genes (e.g., Galimba et al., 2012; Lange et al., 2013). Yeast two-hybrid experiments provide evidence for biochemical interaction amongst the candidate proteins and are typically used as proxy; further studies will be needed to confirm whether these interactions occur *in planta*. Taken together, our results suggest that ThtAP3s can form homodimers, in addition to the AP3-PI heterodimers that preferentially populate B-class/E-class tetramers in *Arabidopsis*

(Melzer and Theissen, 2009). Our detection of AP3 binding sites in the promoters of all B-class genes (**Figure 5**) supports auto- and cross-gene regulation, as has been described for MADS box genes more generally (Kaufmann et al., 2009). We therefore propose that these novel protein interactions provide a potential mechanism for the role of *ThtAP3s* in ectopic petaloidy of sepals.

Although gene duplication has long been recognized as an important contributor to the evolution of biological complexity (Ohno, 1970; Force et al., 1999), functional studies showing the fate of duplicated developmental genes are still limited. Here, we analyzed the function of duplicated AP3 orthologs from a basal eudicot and found deep conservation in stamen identity, a novel role in ectopic petaloidy and stigma development, and partial functional redundancy. While additional experiments will be needed to fully dissect the complex genetic interactions uncovered by our work, evidence presented here lends further support to the overarching hypothesis that the duplication of floral organ identity genes contributed to angiosperm diversification via the generation of floral diversity.

AUTHOR CONTRIBUTIONS

KG performed the experiments and wrote the manuscript. JM-G performed the yeast two-hybrid and promoter analysis and assisted with methods, descriptions, and supplementary materials. VDS designed the study, secured the funding, coordinated the experiments, and edited the manuscript.

FUNDING

This work was funded by NSF-IOS 1121669 to VDS. KG was supported by the National Institutes for Health Training Grant T32 EY07031. JM-G was supported by the Society for Developmental Biology Choose Development! Program (NSF-IOS 1239422) and University of Washington US Department of Education Ronald E. McNair Postbaccalaureate Achievement Program, GenOM Project (NIH-5R25HG007153-03) and Department of Biology's Frye-Hotson-Rigg and UW Biology Scholarships.

ACKNOWLEDGMENTS

We thank the UW Biology Greenhouse staff for plant care and Dr. Edwige Moyroud for helpful conversations regarding motif analysis. This manuscript shares content only with KG's doctoral dissertation (Galimba, 2015), and its publication is in accordance with the University of Washington's dissertation publishing policies.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.00334/full#supplementary-material>

REFERENCES

- Airoidi, C. A., and Davies, B. (2012). Gene duplication and the evolution of plant MADS-box transcription factors. *J. Genet. Genomics* 39, 157–165. doi: 10.1016/j.jgg.2012.02.008 doi: 10.1016/j.jgg.2012.02.008
- Arango-Ocampo, C., González, F., Alzate, J. F., and Pabón-Mora, N. (2016). The developmental and genetic bases of apetaly in *Bocconia frutescens* (Chelidoniaceae: Papaveraceae). *EvoDevo* 7:16. doi: 10.1186/s13227-016-0054-6 doi: 10.1186/s13227-016-0054-6
- Bowman, J. L., Smyth, D. R., and Meyerowitz, E. M. (1989). Genes directing flower development in Arabidopsis. *Plant Cell* 1, 37–52. doi: 10.1105/tpc.1.1.37
- Bowman, J. L., Smyth, D. R., and Meyerowitz, E. M. (1991). Genetic interactions among floral homeotic genes of Arabidopsis. *Development* 112, 1–20.
- Causier, B., and Davies, B. (2002). Analysing protein-protein interactions with the yeast two-hybrid system. *Plant Mol. Biol.* 50, 855–870. doi: 10.1023/A:1021214007897
- Coen, E. S., and Meyerowitz, E. M. (1991). The war of the whorls: genetic interactions controlling flower development. *Nature* 353, 31–37. doi: 10.1038/353031a0
- Cronk, Q. C. B., Bateman, R. M., and Hawkins, J. A. (2002). *Developmental Genetics and Plant Evolution*. Boca Raton, FL: CRC Press.
- Davies, B., Egea-Cortines, M., de Andrade Silva, E., Saedler, H., and Sommer, H. (1996). Multiple interactions amongst floral homeotic MADS box proteins. *EMBO J.* 15, 4330–4343.
- Di Stilio, V. S. (2011). Empowering plant evo-devo: virus induced gene silencing validates new and emerging model systems. *Bioessays* 33, 711–718. doi: 10.1002/bies.201100040
- Di Stilio, V. S., Kramer, E. M., and Baum, D. A. (2005). Floral MADS box genes and homeotic gender dimorphism in *Thalictrum dioicum* (Ranunculaceae) - a new model for the study of dioecy. *Plant J.* 41, 755–766. doi: 10.1111/j.1365-313X.2005.02336.x
- Di Stilio, V. S., Kumar, R. A., Oddone, A. M., Tolkin, T. R., Salles, P., and McCarty, K. (2010). Virus-induced gene silencing as a tool for comparative functional studies in *Thalictrum*. *PLoS One* 5:e12064. doi: 10.1371/journal.pone.0012064 doi: 10.1371/journal.pone.0012064
- Di Stilio, V. S., Martin, C., Schulfer, A. F., and Connelly, C. F. (2009). An ortholog of MIXTA-like2 controls epidermal cell shape in flowers of *Thalictrum*. *New Phytol.* 183, 718–728. doi: 10.1111/j.1469-8137.2009.02945.x
- Drea, S., Hileman, L. C., de Martino, G., and Irish, V. F. (2007). Functional analyses of genetic pathways controlling petal specification in poppy. *Development* 134, 4157–4166. doi: 10.1242/dev.013136
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151, 1531–1545.
- Galimba, K. D. (2015). *Duplication and Functional Divergence in the Floral Organ Identity Genes*. Ph.D. thesis, University of Washington, Seattle, WA.
- Galimba, K. D., Tolkin, T. R., Sullivan, A. M., Melzer, R., Theißen, G., and Di Stilio, V. S. (2012). Loss of deeply conserved C-class floral homeotic gene function and C- and E-class protein interaction in a double-flowered ranunculid mutant. *Proc. Natl. Acad. Sci. U.S.A.* 109, E2267–E2275. doi: 10.1073/pnas.1203686109
- Gonçalves, B., Nougé, O., Jabbour, F., Ridet, C., Morin, H., Laufs, P., et al. (2013). An APETALA3 homolog controls both petal identity and floral meristem patterning in *Nigella damascena* L. (Ranunculaceae). *Plant J.* 76, 223–235. doi: 10.1111/tpj.12284
- Heard, J., and Dunn, K. (1995). Symbiotic induction of a MADS-box gene during development of alfalfa root nodules. *Proc. Natl. Acad. Sci. U.S.A.* 92, 5273–5277. doi: 10.1073/pnas.92.12.5273
- Immink, R. G., Tonaco, I. A., de Folter, S., Shchennikova, A., van Dijk, A. D., Buscher-Lange, J., et al. (2009). SEPALLATA3: the “glue” for MADS box transcription factor complex formation. *Genome Biol.* 10:R24. doi: 10.1186/gb-2009-10-2-r24
- Jaramillo, M. A., and Kramer, E. M. (2004). APETALA3 and PISTILLATA homologs exhibit novel expression patterns in the unique perianth of Aristolochia (Aristolochiaceae). *Evol. Dev.* 6, 449–458. doi: 10.1111/j.1525-142X.2004.04053.x
- Jaramillo, M. A., and Kramer, E. M. (2007). Molecular evolution of the petal and stamen identity genes, APETALA3 and PISTILLATA, after petal loss in the Piperales. *Mol. Phylogenet. Evol.* 44, 598–609. doi: 10.1016/j.ympev.2007.03.015
- Kafri, R., Bar-Even, A., and Pilpel, Y. (2005). Transcription control reprogramming in genetic backup circuits. *Nat. Genet.* 37, 295–299. doi: 10.1038/ng1523
- Kafri, R., Levy, M., and Pilpel, Y. (2006). The regulatory utilization of genetic redundancy through responsive backup circuits. *Proc. Natl. Acad. Sci. U.S.A.* 103, 11653–11658. doi: 10.1073/pnas.0604883103
- Kanno, A., Saeki, H., Kameya, T., Saedler, H., and Theissen, G. (2003). Heterotopic expression of class B floral homeotic genes supports a modified ABC model for tulip (*Tulipa gesneriana*). *Plant Mol. Biol.* 52, 831–841. doi: 10.1023/A:1025070827979
- Kaufmann, K., Muiño, J. M., Jauregui, R., Airoidi, C. A., Smaczniak, C., Krajewski, P., et al. (2009). Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the Arabidopsis flower. *PLoS Biol.* 7:e1000090. doi: 10.1371/journal.pbio.1000090
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858. doi: 10.1038/nprot.2015.053
- Kim, S., Yoo, M.-J., Albert, V. A., Farris, J. S., Soltis, P. S., and Soltis, D. E. (2004). Phylogeny and diversification of B-function MADS-box genes in angiosperms: evolutionary and functional implications of a 260-million-year-old duplication. *Am. J. Bot.* 91, 2102–2118. doi: 10.3732/ajb.91.12.2102
- Kramer, E. M., Di Stilio, V. S., and Schluter, P. (2003). Complex patterns of gene duplication in the APETALA3 and PISTILLATA lineages of the Ranunculaceae. *Int. J. Plant Sci.* 164, 1–11. doi: 10.1086/344694
- Kramer, E. M., Dorit, R. L., and Irish, V. F. (1998). Molecular evolution of genes controlling petal and stamen development: duplication and divergence within the APETALA3 and PISTILLATA MADS-box gene lineages. *Genetics* 149, 765–783.
- Kramer, E. M., Holappa, L., Gould, B., Jaramillo, M. A., Setnikov, D., and Santiago, P. M. (2007). Elaboration of B gene function to include the identity of novel floral organs in the lower eudicot *Aquilegia*. *Plant Cell* 19, 750–766. doi: 10.1105/tpc.107.050385
- Krizek, B. A., and Meyerowitz, E. M. (1996). The Arabidopsis homeotic genes APETALA3 and PISTILLATA are sufficient to provide the B class organ identity function. *Development* 122, 11–22.
- Lange, M., Orashakova, S., Lange, S., Melzer, R., Theißen, G., Smyth, D. R., et al. (2013). The seirena B class floral homeotic mutant of California poppy (*Eschscholzia californica*) reveals a function of the enigmatic PI motif in the formation of specific multimeric MADS domain protein complexes. *Plant Cell* 25, 438–453. doi: 10.1105/tpc.112.105809
- LaRue, N. C., Sullivan, A. M., and Stilio, V. S. D. (2013). Functional recapitulation of transitions in sexual systems by homeosis during the evolution of dioecy in *Thalictrum*. *Front. Plant Sci.* 4:487. doi: 10.3389/fpls.2013.00487
- Lenser, T., Theißen, G., and Ditttrich, P. (2009). Developmental robustness by obligate interaction of class B floral homeotic genes and proteins. *PLoS Comput. Biol.* 5:e1000264. doi: 10.1371/journal.pcbi.1000264
- Litt, A., and Irish, V. F. (2003). Duplication and diversification in the APETALA1/FRUITFULL floral homeotic gene lineage: implications for the evolution of floral development. *Genetics* 165, 821–833.
- Litt, A., and Kramer, E. M. (2010). The ABC model and the diversification of floral organ identity. *Semin. Cell Dev. Biol.* 21, 129–137. doi: 10.1016/j.semcdb.2009.11.019
- Liu, Y. L., Schiff, M., and Dinesh-Kumar, S. P. (2002). Virus-induced gene silencing in tomato. *Plant J.* 31, 777–786. doi: 10.1046/j.1365-313X.2002.01394.x
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262
- Maia, V. H., Gitzendanner, M. A., Soltis, P. S., Wong, G. K.-S., Kane Ka-Shu, and Soltis, D. E. (2014). Angiosperm phylogeny based on 18S/26S rDNA sequence data: constructing a large data set using next-generation sequence data. *Int. J. Plant Sci.* 175, 613–650. doi: 10.1086/676675
- Manchado-Rojas, M., Delgado-Benarroch, L., Roca, M. J., Weiss, J., and Egea-Cortines, M. (2012). Quantitative levels of *Deficiens* and *Globosa* during late petal development show a complex transcriptional network topology of B function. *Plant J.* 72, 294–307. doi: 10.1111/j.1365-313X.2012.05080.x

- Mara, C. D., Huang, T. B., and Irish, V. F. (2010). The *Arabidopsis* floral homeotic proteins APETALA3 and PISTILLATA negatively regulate the BANQUO genes implicated in light signaling. *Plant Cell* 22, 690–702. doi: 10.1105/tpc.109.065946
- Mathelier, A., Zhao, X., Zhang, A. W., Parcy, F., Worsley-Hunt, R., Arenillas, D. J., et al. (2014). JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 42, D142–D147. doi: 10.1093/nar/gkt997
- Medina-Rivera, A., Defrance, M., Sand, O., Herrmann, C., Castro-Mondragon, J. A., Delerue, J., et al. (2015). RSAT 2015: regulatory sequence analysis tools. *Nucleic Acids Res.* 43, W50–W56. doi: 10.1093/nar/gkv362
- Melzer, R., Härter, A., Rümpler, F., Kim, S., Soltis, P. S., Soltis, D. E., et al. (2014). DEF- and GLO-like proteins may have lost most of their interaction partners during angiosperm evolution. *Ann. Bot.* 114, 1431–1443. doi: 10.1093/aob/mcu094
- Melzer, R., and Theißen, G. (2009). Reconstitution of ‘floral quartets’ *in vitro* involving class B and class E floral homeotic proteins. *Nucleic Acids Res.* 37, 2723–2736. doi: 10.1093/nar/gkp129
- Minguet, E. G., Segard, S., Charavay, C., and Parcy, F. (2015). MORPHEUS, a webtool for transcription factor binding analysis using position weight matrices with dependency. *PLoS One* 10:e0135586. doi: 10.1371/journal.pone.0135586
- Moon, Y. H., Jung, J. Y., Kang, H. G., and An, G. (1999). Identification of a rice APETALA3 homologue by yeast two-hybrid screening. *Plant Mol. Biol.* 40, 167–177. doi: 10.1023/A:1026429922616
- Ohno, S. (1970). *Evolution by Gene Duplication*. Berlin: Springer-Verlag. doi: 10.1007/978-3-642-86659-3
- Pabón-Mora, N., Suárez-Baron, H., Ambrose, B. A., and González, F. (2015). Flower development and perianth identity candidate genes in the basal angiosperm *Aristolochia fimbriata* (Piperaceae: Aristolochiaceae). *Front. Plant Sci.* 6:1095. doi: 10.3389/fpls.2015.01095
- Pan, Z.-J., Chen, Y.-Y., Du, J.-S., Chen, Y.-Y., Chung, M.-C., Tsai, W.-C., et al. (2014). Flower development of *Phalaenopsis* orchid involves functionally divergent SEPALLATA-like genes. *New Phytol.* 202, 1024–1042. doi: 10.1111/nph.12723
- Pelaz, S., Gustafson-Brown, C., Kohalmi, S. E., Crosby, W. L., and Yanofsky, M. F. (2001). APETALA1 and SEPALLATA3 interact to promote flower development. *Plant J.* 26, 385–394. doi: 10.1046/j.1365-3113.2001.2641042.x
- Perez-Rodriguez, M., Jaffe, F. W., Butelli, E., Glover, B. J., and Martin, C. (2005). Development of three different cell types is associated with the activity of a specific MYB transcription factor in the ventral petal of *Antirrhinum majus* flowers. *Development* 132, 359–370. doi: 10.1242/dev.01584
- Piwarzyk, E., Yang, Y., and Jack, T. (2007). Conserved C-terminal motifs of the Arabidopsis proteins APETALA3 and PISTILLATA are dispensable for floral organ identity function. *Plant Physiol.* 145, 1495–1505. doi: 10.1104/pp.107.105346
- Puranik, S., Acajjaoui, S., Conn, S., Costa, L., Conn, V., Vial, A., et al. (2014). Structural basis for the oligomerization of the MADS domain transcription factor SEPALLATA3 in *Arabidopsis*. *Plant Cell* 26, 3603–3615. doi: 10.1105/tpc.114.127910
- Riechmann, J. L., Krizek, B. A., and Meyerowitz, E. M. (1996). Dimerization specificity of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS. *Proc. Natl. Acad. Sci. U.S.A.* 93, 4793–4798. doi: 10.1073/pnas.93.10.4793
- Schwarz-Sommer, Z., Hue, I., Huijser, P., Flor, P. J., Hansen, R., Tetens, F., et al. (1992). Characterization of the Antirrhinum floral homeotic MADS-box gene *deficiens*: evidence for DNA binding and autoregulation of its persistent expression throughout flower development. *EMBO J.* 11, 251–263.
- Sharma, B., Guo, C., Kong, H., and Kramer, E. M. (2011). Petal-specific subfunctionalization of an APETALA3 paralog in the Ranunculales and its implications for petal evolution. *New Phytol.* 191, 870–883. doi: 10.1111/j.1469-8137.2011.03744.x
- Sharma, B., and Kramer, E. (2013). Sub- and neo-functionalization of APETALA3 paralogs have contributed to the evolution of novel floral organ identity in *Aquilegia* (columbine, Ranunculaceae). *New Phytol.* 197, 949–957. doi: 10.1111/nph.12078
- Sharma, B., and Kramer, E. M. (2017). *Aquilegia* B gene homologs promote petaloidy of the sepals and maintenance of the C domain boundary. *EvoDevo* 8:22. doi: 10.1186/s13227-017-0085-7
- Soltis, D. E., Smith, S. A., Cellinese, N., Wurdack, K. J., Tank, D. C., Brockington, S. F., et al. (2011). Angiosperm phylogeny: 17 genes, 640 taxa. *Am. J. Bot.* 98, 704–730. doi: 10.3732/ajb.1000404
- Soltis, P., and Soltis, D. (2014). “Flower diversity and angiosperm diversification,” in *Flower Development*, eds J. L. Riechmann and F. Wellmer (New York, NY: Springer), 85–102.
- Soukup, S. W. (1974). “Evolution by gene duplication,” in *Teratology*, Vol. 9, ed. S. Ohno (New York, NY: Springer-Verlag), 250–251.
- Soza, V. L., Brunet, J., Liston, A., Salles Smith, P., and Di Stilio, V. S. (2012). Phylogenetic insights into the correlates of dioecy in meadow-rues (*Thalictrum*, Ranunculaceae). *Mol. Phylogenet. Evol.* 63, 180–192. doi: 10.1016/j.ympev.2012.01.009
- Soza, V. L., Snelson, C. D., Hewett-Hazelton, K. D., and Di Stilio, V. S. (2016). Partial redundancy and functional specialization of E-class SEPALLATA genes in an early-diverging eudicot. *Dev. Biol.* 419, 143–155. doi: 10.1016/j.ydbio.2016.07.021
- Stormo, G. D. (2013). Modeling the specificity of protein-DNA interactions. *Quant. Biol.* 1, 115–130. doi: 10.1007/s40484-013-0012-4
- Su, K., Zhao, S., Shan, H., Kong, H., Lu, W., Theissen, G., et al. (2008). The MIK region rather than the C-terminal domain of AP3-like class B floral homeotic proteins determines functional specificity in the development and evolution of petals. *New Phytol.* 178, 544–558. doi: 10.1111/j.1469-8137.2008.02382.x
- Theissen, G., Kim, J. T., and Saedler, H. (1996). Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes. *J. Mol. Evol.* 43, 484–516. doi: 10.1007/BF02337521
- Theißen, G., and Saedler, H. (2001). Plant biology: floral quartets. *Nature* 409, 469–471. doi: 10.1038/35054172
- Tilly, J. J., Allen, D. W., and Jack, T. (1998). The CARG boxes in the promoter of the Arabidopsis floral organ identity gene APETALA3 mediate diverse regulatory effects. *Development* 125, 1647–1657.
- Tröbner, W., Ramirez, L., Motte, P., Hue, I., Huijser, P., Lönning, W. E., et al. (1992). GLOBOSA: a homeotic gene which interacts with DEFICIENS in the control of *Antirrhinum* floral organogenesis. *EMBO J.* 11, 4693–4704.
- van Tunen, A. J., Eikelboom, W., and Angenent, G. C. (1993). Floral organogenesis in *Tulipa*. *Flower. Newsl.* 16, 33–37.
- Vandenbussche, M., Zethof, J., Royaert, S., Weterings, K., and Gerats, T. (2004). The duplicated B-class heterodimer model: whorl-specific effects and complex genetic interactions in *Petunia hybrida* flower development. *Plant Cell* 16, 741–754. doi: 10.1105/tpc.019166
- Wang, P., Liao, H., Zhang, W., Yu, X., Zhang, R., Shan, H., et al. (2015). Flexibility in the structure of spiral flowers and its underlying mechanisms. *Nat. Plants* 2:15188. doi: 10.1038/nplants.2015.188
- Wuest, S. E., O’Maoileidigh, D. S., Rae, L., Kwasniewska, K., Raganelli, A., Hanczaryk, K., et al. (2012). Molecular basis for the specification of floral organs by APETALA3 and PISTILLATA. *Proc. Natl. Acad. Sci. U.S.A.* 109, 13452–13457. doi: 10.1073/pnas.1207075109
- Zahn, L. M., Leebens-Mack, J., dePamphilis, C. W., Ma, H., and Theissen, G. (2005). To B or not to B a flower: the role of DEFICIENS and GLOBOSA orthologs in the evolution of the angiosperms. *J. Hered.* 96, 225–240. doi: 10.1093/jhered/esi033
- Zhang, R., Guo, C., Zhang, W., Wang, P., Li, L., Duan, X., et al. (2013). Disruption of the petal identity gene APETALA3-3 is highly correlated with loss of petals within the buttercup family (Ranunculaceae). *Proc. Natl. Acad. Sci. U.S.A.* 110, 5074–5079. doi: 10.1073/pnas.1219690110

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 © 2018 Galimba, Martínez-Gómez and Di Stilio. 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 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.



Unraveling the Developmental and Genetic Mechanisms Underpinning Floral Architecture in Proteaceae

Catherine Damerval^{1*}, Hélène Citerne¹, Natalia Conde e Silva¹, Yves Deveaux¹, Etienne Delannoy², Johann Joets¹, Franck Simonnet^{1,3}, Yannick Staedler⁴, Jürg Schönenberger⁴, Jennifer Yansouni², Martine Le Guilloux¹, Hervé Sauquet^{3,5} and Sophie Nadot^{3*}

OPEN ACCESS

Edited by:

Zhongchi Liu,
University of Maryland, College Park,
United States

Reviewed by:

Oriane Hidalgo,
Royal Botanic Gardens, Kew,
United Kingdom
Hongzhi Kong,
Institute of Botany (CAS), China
Verónica S. Di Stilio,
University of Washington,
United States

*Correspondence:

Catherine Damerval
catherine.damerval@u-psud.fr
Sophie Nadot
sophie.nadot@u-psud.fr

Specialty section:

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

Received: 12 June 2018

Accepted: 08 January 2019

Published: 25 January 2019

Citation:

Damerval C, Citerne H,
Conde e Silva N, Deveaux Y,
Delannoy E, Joets J, Simonnet F,
Staedler Y, Schönenberger J,
Yansouni J, Le Guilloux M, Sauquet H
and Nadot S (2019) Unraveling
the Developmental and Genetic
Mechanisms Underpinning Floral
Architecture in Proteaceae.
Front. Plant Sci. 10:18.
doi: 10.3389/fpls.2019.00018

¹ GQE-Le Moulon, INRA, Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, Gif-sur-Yvette, France, ² Institute of Plant Sciences Paris-Saclay, CNRS, INRA, Universités Paris Diderot, Paris-Sud, Evry, Paris-Saclay, Gif-sur-Yvette, France, ³ Ecologie Systématique Evolution, AgroParisTech, CNRS, Univ. Paris-Sud, Université Paris-Saclay, Orsay, France, ⁴ Department of Botany and Biodiversity Research, University of Vienna, Vienna, Austria, ⁵ National Herbarium of New South Wales (NSW), Royal Botanic Gardens and Domain Trust, Sydney, NSW, Australia

Proteaceae are a basal eudicot family with a highly conserved floral groundplan but which displays considerable variation in other aspects of floral and inflorescence morphology. Their morphological diversity and phylogenetic position make them good candidates for understanding the evolution of floral architecture, in particular the question of the homology of the undifferentiated perianth with the differentiated perianth of core eudicots, and the mechanisms underlying the repeated evolution of zygomorphy. In this paper, we combine a morphological approach to explore floral ontogenesis and a transcriptomic approach to access the genes involved in floral organ identity and development, focusing on *Grevillea juniperina*, a species from subfamily Grevilleoideae. We present developmental data for *Grevillea juniperina* and three additional species that differ in their floral symmetry using stereomicroscopy, SEM and High Resolution X-Ray Computed Tomography. We find that the adnation of stamens to tepals takes place at early developmental stages, and that the establishment of bilateral symmetry coincides with the asymmetrical growth of the single carpel. To set a framework for understanding the genetic basis of floral development in Proteaceae, we generated and annotated *de novo* a reference leaf/flower transcriptome from *Grevillea juniperina*. We found *Grevillea* homologs of all lineages of MADS-box genes involved in floral organ identity. Using *Arabidopsis thaliana* gene expression data as a reference, we found homologs of other genes involved in floral development in the transcriptome of *G. juniperina*. We also found at least 21 class I and class II TCP genes, a gene family involved in the regulation of growth processes, including floral symmetry. The expression patterns of a set of floral genes obtained from the transcriptome were characterized during floral development to assess their organ specificity and asymmetry of expression.

Keywords: Proteaceae, flower, development, floral symmetry, High Resolution X-Ray Computed Tomography, transcriptome, MADS-box genes, TCP genes

INTRODUCTION

Proteaceae are a family of woody plants comprising approximately 1700 species in 81 genera, distributed mainly in the Southern Hemisphere, with two main centers of diversity, one in Australia and the other in South Africa. The hypogynous flowers almost invariably consist of four tepals (rarely 3 or 5), four stamens (rarely 3 or 5) opposite the tepals and with filaments that are adnate to the tepals (rarely free), and a single carpel with marginal placentation (Weston, 2007). Although this floral groundplan is highly conserved, the family displays considerable variation in other aspects of floral and inflorescence morphology. The inflorescence is basically a raceme but with various degrees of compaction. In subfamily Grevilleoideae, all but two genera are characterized by compound inflorescences consisting of racemes of flower pairs described as two-flowered short shoots sharing a common bract (Douglas and Tucker, 1996a). Early floral development has been described in detail in several species of Grevilleoideae, with emphasis on the ontogenic origin of the flower pair (Douglas and Tucker, 1996a) and carpel orientation (Douglas and Tucker, 1996b). In addition, Proteaceae have the highest number of transitions in perianth symmetry across all angiosperms (Reyes et al., 2016), with at least 10 transitions from actinomorphy (radial symmetry) to zygomorphy (bilateral symmetry) inferred throughout the family, and at least four reversals (Citerne et al., 2017). The developmental stage at which bilateral symmetry becomes visible varies across angiosperm species with a zygomorphic perianth; zygomorphy can be present from the very first stages, when the first floral organs are initiated, or can appear after all organs have been initiated and have begun to differentiate (Endress, 1999; Tucker, 1999). As in many taxa with a monomerous gynoecium, the single plicate carpel in Proteaceae becomes bilaterally symmetrical when the cleft starts to form, making the flower zygomorphic at the gynoecium level (Douglas and Tucker, 1996b; Sokoloff et al., 2017). Furthermore, in Proteaceae and more specifically in the subfamily Grevilleoideae, the high diversity of carpel orientation (dorso-ventral or oblique) adds further complexity when defining the orientation of the plane of symmetry relative to the common axis of the flower pair (Douglas and Tucker, 1996b).

The molecular bases of floral symmetry have been investigated in depth in *Antirrhinum majus* (Plantaginaceae). The *CYCLOIDEA* (CYC) gene is essential for the asymmetric development of the flower in the petal and stamen whorls (Luo et al., 1996, 1999). CYC belongs to a plant-specific transcription factor gene family, the TCP family, which is characterized by a non-canonical basic helix-loop-helix domain (the TCP domain, Cubas et al., 1999). TCP genes have been reported to be involved in various cell proliferation and growth processes that control organ development and shape (Huang and Irish, 2015; Nicolas and Cubas, 2016). In *A. majus*, CYC expression varies with the identity of the floral organ. Organ identity is classically determined by the combined action of four classes of genes, referred to as the A-, B-, C-, and E-classes (Coen and Meyerowitz, 1991; Causier et al., 2010). Petals are determined by the combined action of A- and B-class genes, stamens by B- and C-class genes,

sepals by A-class genes only, and the carpel by C-class genes only. E-class genes act as obligate partners for all floral organ identity genes. The model is generally conserved across angiosperms with the notable exception of the A-class, whose function as defined in *Arabidopsis thaliana* appears to be limited to close relatives in the Brassicaceae. With the exception of one A function gene in *A. thaliana*, namely *APETALA2*, these floral organ identity genes belong to the MADS-box transcription factor gene family (Theißen et al., 1996). B- and C-class genes may play a role in maintaining expression of CYC in the second and third floral whorls in *A. majus* (Clark and Coen, 2002).

Asymmetric expression patterns of *CYCLOIDEA* homologs (hereafter CYC-like genes) have been correlated with the independent evolution of zygomorphy in many monocot and eudicot clades (e. g., Hileman, 2014; Spencer and Kim, 2018). Functional studies have confirmed the role of CYC-like genes in the asymmetric development of floral organs in distantly related core eudicot species such as *Lotus japonicus*, *Iberis amara*, a hybrid of *Gerbera*, and in the monocot *Oryza sativa* (Feng et al., 2006; Busch and Zachgo, 2007; Broholm et al., 2008; Yuan et al., 2009). In orchids, however, no TCP gene has been found to date to be involved in zygomorphy. In this large monocot family, duplication and subfunctionalization in one of the two B-class gene lineages account for the inter- and intra-whorl tepal differentiation that generates zygomorphy (reviewed in Mondragon-Palomino, 2013). B-class genes have been found to be implicated in the independent evolution of zygomorphy in other monocot taxa. For instance, in *Commelina communis* (Commelinaceae), it has been suggested that in addition to the asymmetric expression of a CYC-like gene, zygomorphy is mediated by the asymmetric expression of a B-class gene in the inner tepal whorl associated with a change in ventral organ identity (from petal to sepal) (Preston and Hileman, 2012). In maize, the lack of expression of the B-class genes in the dorsal domain coincides with the absence of initiation of the dorsal lodicule (Bartlett et al., 2015).

Proteaceae, with their morphological diversity and their phylogenetic position within the basal eudicots, are good candidates for understanding the evolution of floral architecture, particularly regarding the homology of the undifferentiated perianth with either the calyx or the corolla of a differentiated perianth, and the repeated evolution of zygomorphy. In addition, the origin of the tetramerous perianth remains unclear. A recent study of floral trait evolution across angiosperms suggested that the ancestral perianth of eudicots may have been dimerous and undifferentiated (Sauquet et al., 2017), in which case the tetramerous perianth of Proteaceae could be of dimerous origin inherited from the ancestor of all eudicots. Yet, Proteaceae are relatively understudied from the point of view of developmental genetics. CYC-like genes have recently been characterized in Proteaceae. In two zygomorphic species of *Grevillea*, expression of the same *ProtCYC* gene was found to be asymmetric during the late stages of floral development (Citerne et al., 2017).

This paper combines developmental and transcriptomic data in *Grevillea juniperina* from subfamily Grevilleoideae. We compared early floral ontogenesis in *G. juniperina* with three other species presenting similar floral features but differing

in floral symmetry, using a range of microscopy techniques (including High Resolution X-Ray Computed Tomography). We assembled *de novo* a transcriptome of *Grevillea juniperina* from RNA-seq of leaf and flower tissues. We searched this transcriptome for homologs of genes known to be involved in floral development in *Arabidopsis*, as well as focusing on the entire TCP and A-, B-, C-, E-class MADS-box gene families. We characterized the expression pattern during floral development of selected genes in *G. juniperina*.

MATERIALS AND METHODS

Plant Material

Plant material (inflorescences at various pre-anthetic stages) was collected from four species of Proteaceae belonging to tribe Embothrieae of subfamily Grevilleoideae (Weston and Barker, 2006) and endemic to Australia: *Alloxylon flammeum* P. H. Weston & Crisp (subtribe Embothriinae), *Stenocarpus davallioides* D. Foreman & B. Hyland (subtribe Stenocarpinae), *Grevillea juniperina* R. Br. and *Grevillea petrophiloides* Meisn. (subtribe Hakeinae). The first three species have zygomorphic flowers while the latter has actinomorphic flowers (**Figure 1**). Detailed descriptions and illustrations of these species are available in the Flora of Australia online¹. A dissection of the flower of *Grevillea juniperina*, showing all floral organs, is shown in Citerne et al. (2017).

Material for *Alloxylon flammeum* and *Stenocarpus davallioides* was sampled from trees cultivated at the Royal Botanic Garden Sydney (under accessions S2003-0047 and S1986-2016, respectively; voucher specimens were made (P.H.Weston 3425 and P.H.Weston 3426) and deposited at The National Herbarium of New South Wales (NSW)) and fixed in FAA (85 mL 55% ethanol, 5 mL glacial acetic acid, 10 mL formaldehyde). Material for *Grevillea petrophiloides* was sampled from trees cultivated in Longueville, NSW (voucher P.H.Weston 3582: High Resolution X-ray Computed Tomography) and Oakdale, NSW (voucher P.M.Olde 18/01: light microscopy) and fixed in FAA and 70% alcohol, respectively. Five plants of *Grevillea juniperina* were obtained from the garden center Truffaut and planted in the Parc Botanique de Launay (Orsay, France). These plants provided fresh material for the development analysis and for RNA extraction.

Development Analysis

Light Microscopy and Scanning Electron Microscopy

Light microscopy was carried out on a Leica M125 stereomicroscope. For scanning electron microscopy, developing inflorescences and dissected tissues were dehydrated through an ethanol-acetone series, critical-point dried and sputter coated with gold. Imaging was carried out on a JEOL JSM 6390 Scanning Electron Microscope at 10 kV.

High Resolution X-Ray Computed Tomography (HRXCT) and 3D Reconstruction

For HRXCT, developing inflorescences of all four species were treated with a solution of 1% (w/v) phosphotungstic acid in FAA for 2 months (thereby ensuring saturation of the material with phosphotungstic acid), following the protocol of Staedler et al. (2013). The scans were performed on a MicroXCT-200 system (Zeiss Microscopy). X-Ray detection was performed via scintillator crystals (Zeiss Microscopy, in-house). The X-ray source was a Hamamatsu L9421-02 90 kV Microfocus X-Ray source. XMReconstructor 8.1.6599 (Zeiss Microscopy) was used to perform the 3D reconstruction from the scanning data. The AMIRA-based XM3DViewer 1.1.6 (XRadia Inc.) was used to visualize the scans. The reconstructed 3D data were then exported with XMController 8.1.6599 as a series of pictures in tiff format, typically ca. one to two thousand per sample. Scanning parameters are summarized in **Supplementary Table S1**.

Tissue Collection and RNA Extraction

Young leaves and flower buds of *Grevillea juniperina* were collected and processed as in Citerne et al. (2017). Total RNA was extracted following a 2x CTAB-based extraction buffer and lithium chloride precipitation method (Smart and Roden, 2010) and total RNA was treated with DNase (Ambion) following the manufacturer's instructions. For RNA-seq analysis, all tissues were harvested on the same plant: young leaves and floral buds of three size ranges (1–2 mm, 2–5 mm, over 6 mm long), and processed individually.

For qPCR analyses, floral buds from different plants were combined to constitute three biological replicates for flower dissections. Dissections were carried out on freshly harvested 6–7 mm buds already exhibiting bilateral symmetry under a stereomicroscope (6–10 buds from two–three plants for each biological replicate). Three tissue types (dorsal tepals with adnate stamens, ventral tepals with adnate stamens and gynoecium) were separately frozen in liquid nitrogen.

Reference Transcriptome Sequencing

Total RNAs of each tissue were checked for their integrity on a RNA_Nano chip, using an Agilent 2100 bioanalyzer (Agilent Technologies, Waldbronn, Germany), then pooled in equal amounts. The RNA-seq procedure was carried out at the Institute of Plant Sciences Paris-Saclay (IPS2, Saclay, France) on an IG-CNS Illumina HiSeq2000 platform. The library was constructed with the TruSeq stranded mRNA library Prep kit (Illumina®, CA, United States) with a sizing of 260 bp, and then sequenced in paired-end (PE) and a read length of 100 bases on a single lane. After quality trimming (removing the adapter, trimming of bases with a Q score <20 and removing pairs with at least one read <30 bases), 186,060,539 pairs of reads were generated. The raw data are available at SRA under the accession SRP141177.

¹ <http://www.environment.gov.au/science/abrs/online-resources/flora-of-australia-online>

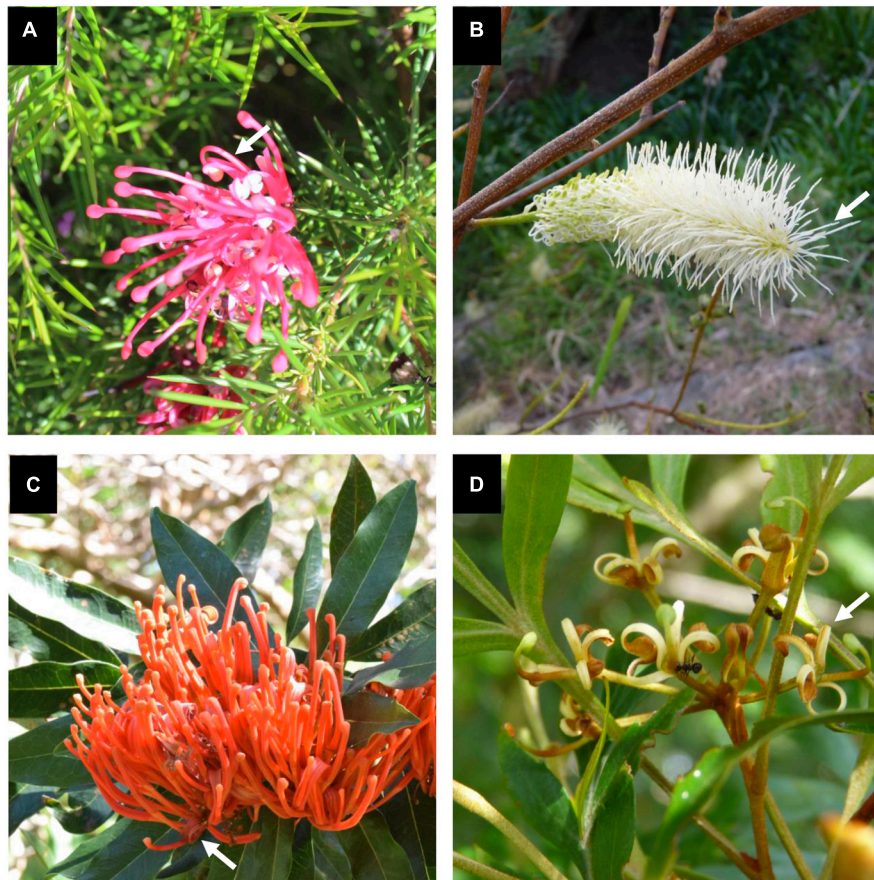


FIGURE 1 | Inflorescences of the study species. All species have tetramerous flowers with stamens completely adnate to tepals. **(A)** Inflorescence of *Grevillea juniperina* with zygomorphic flowers at anthesis (arrow pointing to an open flower displaying curved tepals and style). **(B)** Inflorescence of *Grevillea petrophiloides* with actinomorphic flowers at anthesis (arrow pointing to an open flower displaying straight tepals and style). **(C)** Inflorescence of *Alloxylon flammeum* with zygomorphic flowers at anthesis (arrow pointing to an open flower displaying curved tepals and style). **(D)** Inflorescence of *Stenocarpus davallioides* with zygomorphic flowers at anthesis (arrow pointing to an open flower displaying curved tepals and style). Photographs S. Nadot **(A)** and H. Sauquet **(B–D)**.

Transcriptome Assembly

The *de novo* assembly of the transcriptome was performed following Roberts and Roalson (2017). Six assemblies were generated with Trinity v2.3.2 (Haas et al., 2013, with default parameters except `-SS_lib_typeRF` and `-min_kmer_cov 3`) and Velvet-Oases v1.2.09 (Schulz et al., 2012, kmers 25, 35, 45, 55, and 65) and combined with the tr2aacs.pl script (v 2014.05.15) of the Evidential Genes suite (Nakasugi et al., 2014). Only contigs classified as “main” were kept. The quality of the final assembly was assessed with BUSCO v3.0.2 (Waterhouse et al., 2018) using hmmer 3.1b2 against the embryophyta_odb9 dataset.

Homology Search and Functional Annotation

Predicted protein sequences from the *Grevillea juniperina* assembled transcriptome were annotated with BLASTP best hit (Camacho et al., 2009, *e*-value cutoff $1e-3$) against a local version of the *Arabidopsis thaliana* protein database (TAIR 10). From BLASTP results, a gene ontology (GO) annotation was generated and extended through merging with

InterProScan results using default parameters (Jones et al., 2014).

Overall homology of the *G. juniperina* proteome with other eudicot proteomes was assessed using OrthoVenn² (Wang et al., 2015), selecting the proteomes of *Arabidopsis thaliana* and *Vitis vinifera*, and uploading the proteome of *Nelumbo nucifera* from NCBI (GCF_000365185.1_Chinese_Lotus_1.1_protein.faa.gz).

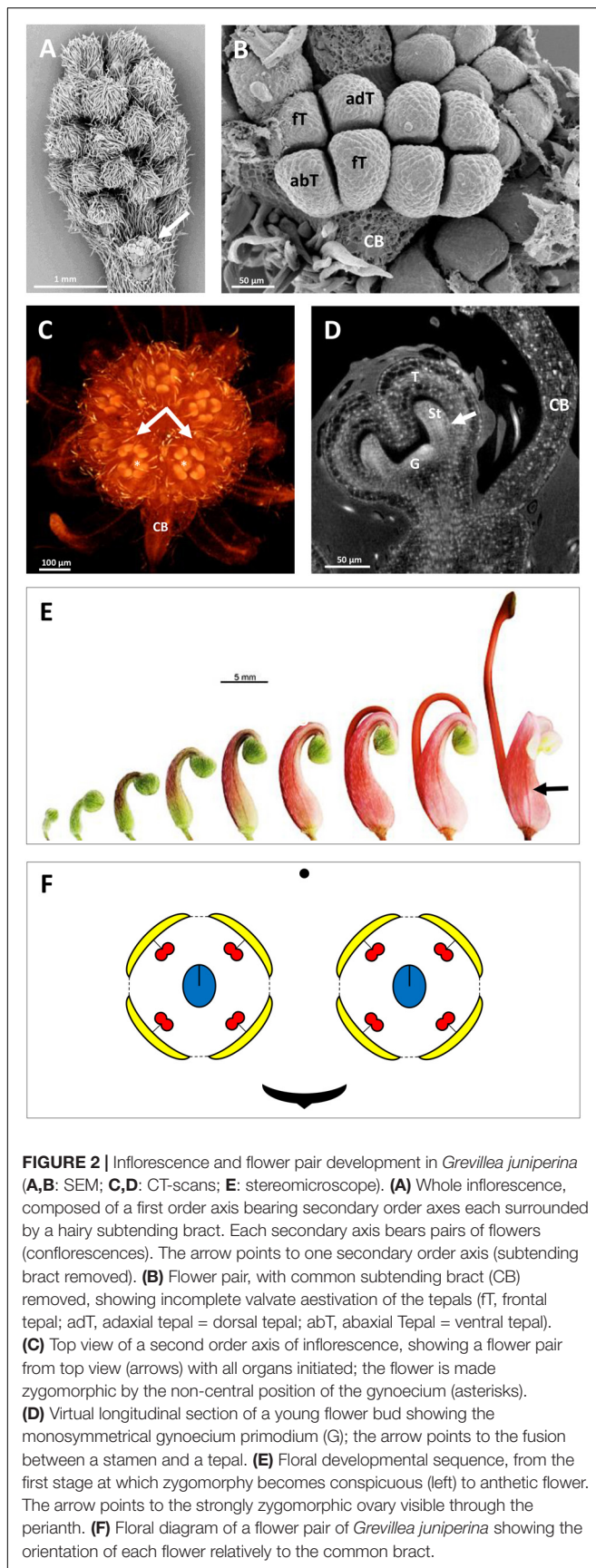
Floral Gene Mining and Phylogenetic Analyses

Phylogeny of TCP Genes

All *Grevillea* contigs identified as TCP genes were extracted from the *G. juniperina* transcriptome annotated against the proteome of *Arabidopsis thaliana* by BLAST homology search. Sequence alignment was done with MUSCLE³ (Edgar, 2004) based on predicted amino acid translations, then manually refined. Phylogenetic reconstruction was done using

²<http://www.bioinfogenome.net/OrthoVenn/>

³<https://www.ebi.ac.uk/Tools/msa/muscle/>



maximum likelihood as implemented in PhyML (Guindon et al., 2010), using the GTR + Γ model [substitution model selected by AIC using the SMS option in PhyML (Lefort et al., 2017)] and the NNI method of tree optimization. Branch support was calculated with the aLRT SH-like method.

Phylogeny of ABCE MADS-Box Coding Sequences

In *Arabidopsis*, MADS-box genes with floral organ identity function are *APETALA1* (*AP1*) (A-class), *APETALA3* (*AP3*) and *PISTILLATA* (*PI*) (B-class), *AGAMOUS* (*AG*) (C-class) and *SEPALLATA 1/2/4* (*SEP1/2/4*) and *SEPALLATA3* (*SEP3*) (E-class). We searched for homologs of these MADS-box genes in the *G. juniperina* transcriptome. Sixteen contigs were retrieved from the transcriptome following annotation against the *Arabidopsis thaliana* proteome. Sequences from transcriptomic data were translated, and seven full length protein sequences were used for subsequent analyses. MADS-box coding sequences (CDS) from *G. juniperina* were aligned with known floral MADS-box CDS of *Arabidopsis thaliana*⁴, *Nelumbo nucifera*⁵, *Aquilegia caerulea*⁶, and *Vitis vinifera*⁶, using MUSCLE based on predicted amino acid translations⁷ (protein IDs in **Supplementary Table S2**). Phylogenetic reconstruction was done using maximum likelihood as implemented in PhyML (Guindon et al., 2010), using the GTR + Γ + I model [substitution model selected by AIC using the SMS option in PhyML (Lefort et al., 2017)] and the NNI method of tree optimization. Branch support was calculated with the aLRT SH-like method.

Homologs of Floral Genes Co-expressed With ABCE MADS-Box Genes in *Grevillea juniperina*

We searched for the genes whose expression is correlated with that of the MADS-box genes implicated in the ABCE model in *A. thaliana*, excluding *SEP4* because of its low expression specificity to reproductive tissues. The search was done with the Expression Angler program using the AtGenExpress - Plus Extended tissue Compendium data set (Toufighi et al., 2005). The Pearson correlation coefficient (*r*-value) cut-off was fixed at 0.5 or 0.75, and a set of unique AGI (*Arabidopsis* Genome Initiative) gene identifiers (AGI-ID) was created. For each AGI-ID, the frequency of occurrence was calculated for the seven *A. thaliana* MADS-box genes used as baits and the best *r*-value reported (**Supplementary Table S3**). Then we searched for homologs of these genes in the *G. juniperina* transcriptome annotated against the *A. thaliana* proteome. A new dataset of the best matching AGI-ID in the *G. juniperina* transcriptome was obtained and the corresponding numbers of *G. juniperina* contigs were reported (**Supplementary Table S3**). A list of the 25 best correlated genes was also generated for each of the seven MADS-box genes used as baits and compared to the *G. juniperina* transcriptome.

⁴<https://www.arabidopsis.org/>

⁵<https://www.ncbi.nlm.nih.gov/>

⁶<https://phytozome.jgi.doe.gov/>

⁷<http://translatrix.co.uk/>

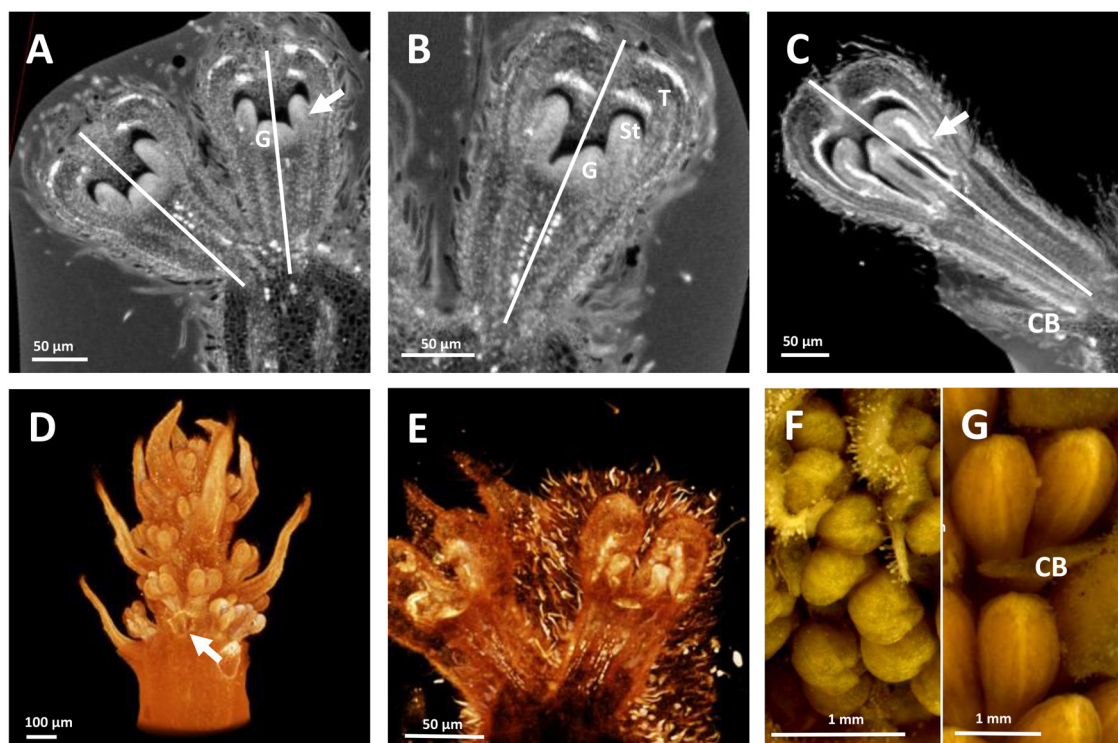


FIGURE 3 | HRXCT images of developing flowers and flower pairs in *Stenocarpus davallioides* (A–C), *Alloxyylon flammeum* (D,E) and HRXCT and stereomicroscope images of inflorescence and flower pairs in *Grevillea petrophiloides* (F,G). (A) Virtual longitudinal section of a flower pair (lateral view), with slight monosymmetry of the gynoecium (G) with regard to the vertical floral axis, visible on the right; the arrow points to the fusion between a stamen and a tepal. (B) Virtual longitudinal section of a flower (front view), showing the monosymmetry of the gynoecium in the vertical axis, while the tepals (T) and adnate stamens (St) are identical. (C) Virtual longitudinal section of a flower at a later stage (lateral view), showing slight asymmetry of the tepals and adnate stamens (the arrow points to the fused zone between a stamen and a tepal) with respect to the vertical axis (CB, common bract). (D) Lateral view of a whole inflorescence, with the common subtending bract of a flower pair removed (arrow). (E) Lateral view of a flower, zygomorphy is not yet apparent. (F,G) Lateral view of flower pairs at two developmental stages (CB, common bract).

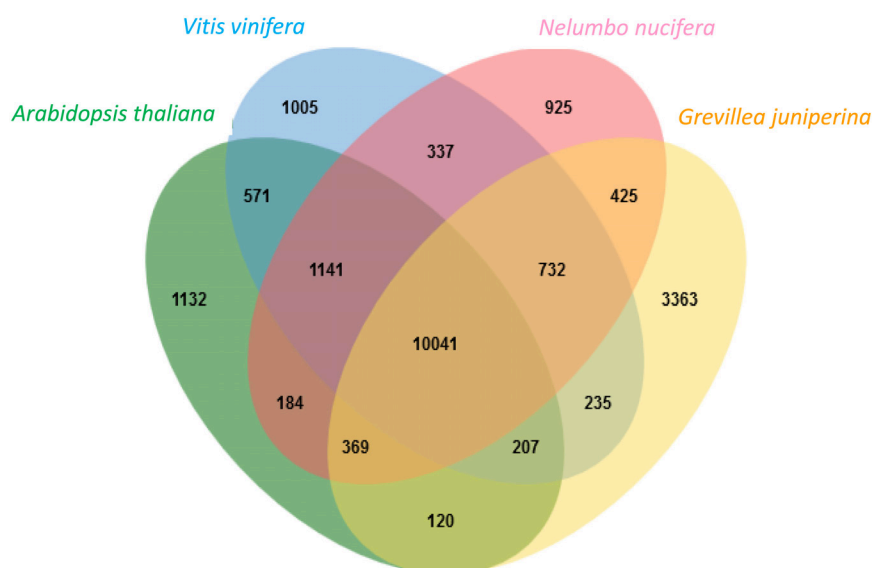


FIGURE 4 | Distribution of clusters (orthologous groups) between the *Arabidopsis thaliana*, *Vitis vinifera*, *Nelumbo nucifera* proteomes and the *Grevillea juniperina* leaf/flower proteome.

Quantitative RT-PCR Analyses of Floral Gene Expression

Thirteen *G. juniperina* contigs were selected based on their homology with genes expressed during floral development in *Arabidopsis thaliana*. They were renamed according to homology as follows: *AGAMOUS* (*GjuAG2*), *PISTILLATA* (*GjuPI2*), *SEPALLATA* (*GjuSEP1* and *GjuSEP3*), *APETALA2* (*GjuAP2*), *APETALA1* (*GjuAPI*), *WUSCHEL* (*GjuWUS*), *AHOOTMERISTEMLESS* (*GjuSTM*), *CUP-SHAPED COTYLEDON2* (*GjuCUC2*), *SPATULA* (*GjuSPT*), *BEL1* (*GjuBEL1*), *TOUSLED* (*GjuTSL*), and *CRABSCRAW* (*GjuCRC*). In addition, we included the *ProtCYC* genes (*GjuCYC1* and *GjuCYC2*) that were found to be expressed in flower buds by Citerne et al. (2017). Expression of these 15 genes was investigated in dissections of floral organs of pre-anthetic flowers (as described above).

Primer pairs were designed from the contig sequences (Supplementary Table S4). The *GjuPI2* gene corresponds to a contig closely related to the full length *GjuPI* used in the phylogenetic reconstruction, but lacking C-terminal region. Unfortunately, we were unable to define an appropriate pair of specific primers for *GjuAPI* from the available sequence. qRT-PCR reactions were performed using the Bio-Rad CFX384 touch (Bio-Rad, France) and the SYBR Premix Ex Taq (Tli RNaseH Plus) (Takara, Ozyme, France) following the supplier's instructions. Three technical replicates were done for each of the three biological replicates of the floral dissections. A dilution series of the pooled cDNAs was used as a standard curve to validate the primer pairs and estimate the starting quantities (arbitrary units). Gene expression was normalized with the mean of the three reference genes *ACT8*, *TUB7*, and *FBA1* coding for actin, tubulin and fructose bisphosphate aldolase, respectively. The normalized expression values are provided in Supplementary Table S5. Differential expression for each gene was tested with a Two-way ANOVA (tissue and replicate effects) and a Tukey test for pairwise comparisons. *P*-values were adjusted for multiple testing with a BH correction. Adjusted *p*-values lower than 0.05 were considered as significant.

RESULTS

Floral Development and Establishment of Zygomorphy

The developing flowers in *Grevillea juniperina* (zygomorphic flowers at anthesis, Figure 1A, arrow) display a perianth that is actinomorphic until the tepals are formed and begin to elongate (Figures 2A–F), as in the actinomorphic species *Grevillea petrophiloides* (Figure 1B, distal part of the inflorescence, and Figures 3E,G). In *G. juniperina*, the virtual longitudinal section obtained with HRXCT shows that the carpel develops asymmetrically relative to the center of the floral bud (Figure 2D), even though the other organ whorls (tepals and stamens) are radially symmetrical. The monosymmetrical shape of the developing carpel suggests asymmetrical growth of the primordium. It can be noted that the incomplete valvate aestivation in *G. juniperina* (Figure 2B)

makes the young flower bud slightly disymmetric rather than strictly actinomorphic. The base of the tepals and stamens is already slightly fused, indicating that the adnation of stamens to tepals occurs early during development. Early development is similar in the two other species sampled in this study, *Alloxylon flammeum* and *Stenocarpus davallioides* (both zygomorphic at anthesis; Figures 1C,D, arrows), which have flowers with an actinomorphic perianth at early stages and a gynoeceum that develops asymmetrically relative to the center of the floral bud (Figures 3A–E). In *Grevillea petrophiloides*, the style elongates faster than the tepals before anthesis, and forces its way through the two dorsal tepals (visible in the proximal part of the inflorescence, Figure 1B), but in mature flowers, the style and the tepals are straight, contrary to the three other species.

Assembly and Annotation of the Reference Transcriptome

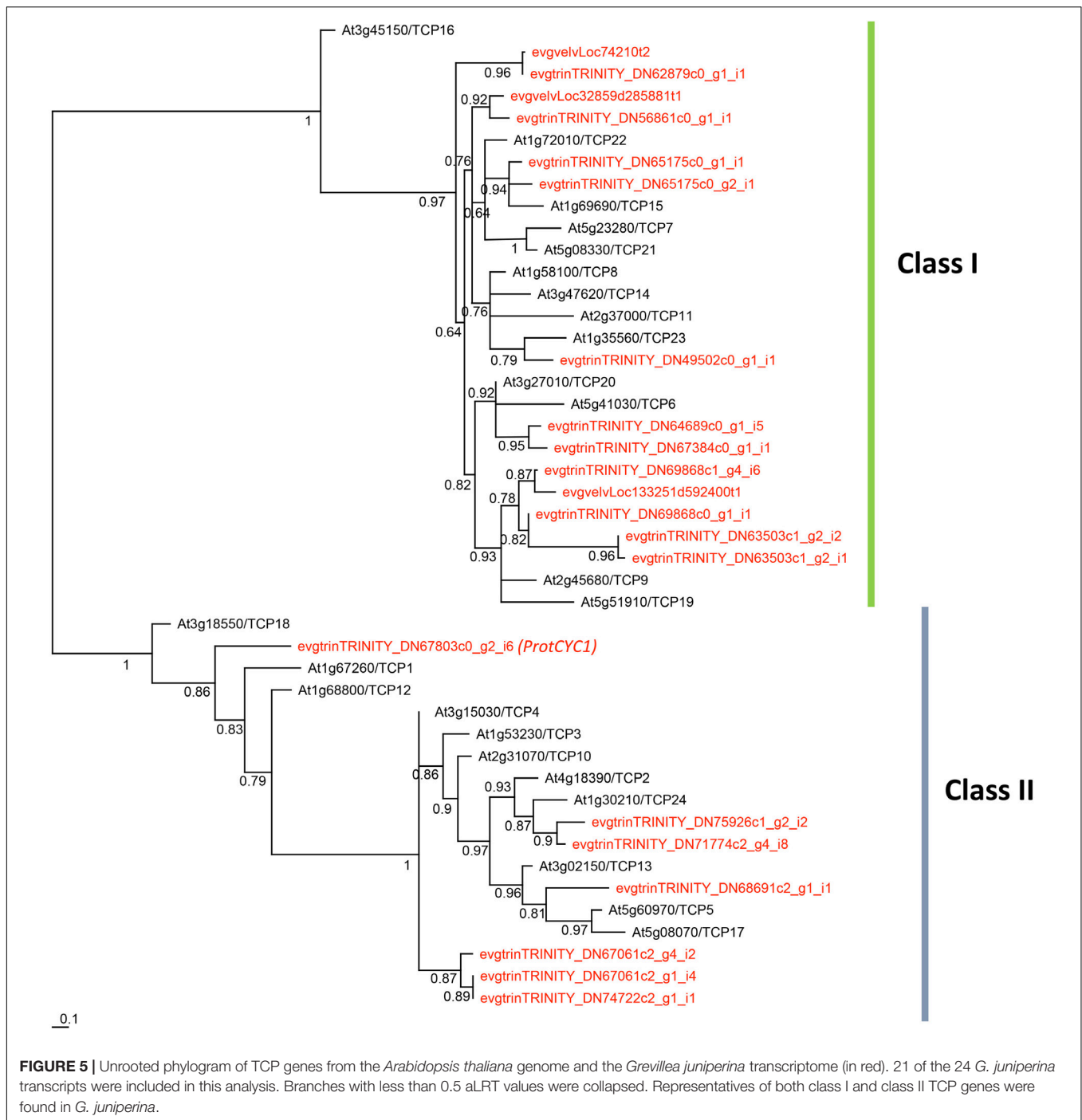
67,004 contigs were obtained from the *de novo* assembly of the transcriptome of *G. juniperina*, with lengths ranging from 180 to 9,937 bases for a mean of 877.8, a N50 of 1,491 and 59,306,330 assembled bases. The result of the BUSCO analysis was C: 84.8% [S: 74.9%, D: 9.9%], F: 8.4%, M: 6.8%, n: 1,440 demonstrating the good quality of this assembly.

InterProScan identified 30,063 contigs with at least one protein domain. Among these, 14,488 contigs could be annotated with one category (or more) of the Gene Ontology (GO). We focused on transcription factors, selecting all contigs annotated as GO:0003700 and analyzed their PFAM domains. We found 240 contigs with such transcription domain annotation (1.6% of the GO annotated contigs), most of which corresponded to AP2 (29%), WRKY DNA-binding (23%) and bZip (12%) domains (Supplementary Figure S1).

37,842 contigs had homology with 13,946 *Arabidopsis* proteins, while 37,817 had a GO annotation. The two best-represented categories in Biological Process were “other cellular processes” (26%) and “other metabolic processes” (22%); in Cellular Components, the four best-represented categories were “other cytoplasmic components” (18%), “other intracellular processes” (14%), “nucleus” (15%) and “other membranes” (13%); among Molecular Functions, the prominent categories were “protein binding” (20%), “other binding” (14%), “other enzyme activity” (13%), and “transferase activity” (11%) (Supplementary Figure S2). We found 937 *G. juniperina* contigs with a functional annotation in one or more categories identified as GO: 0009908 (flower development) and its children categories. Beyond general flower development, the most represented categories pertained to stamen (20%) and ovule and carpel (16%) development (Supplementary Figure S3).

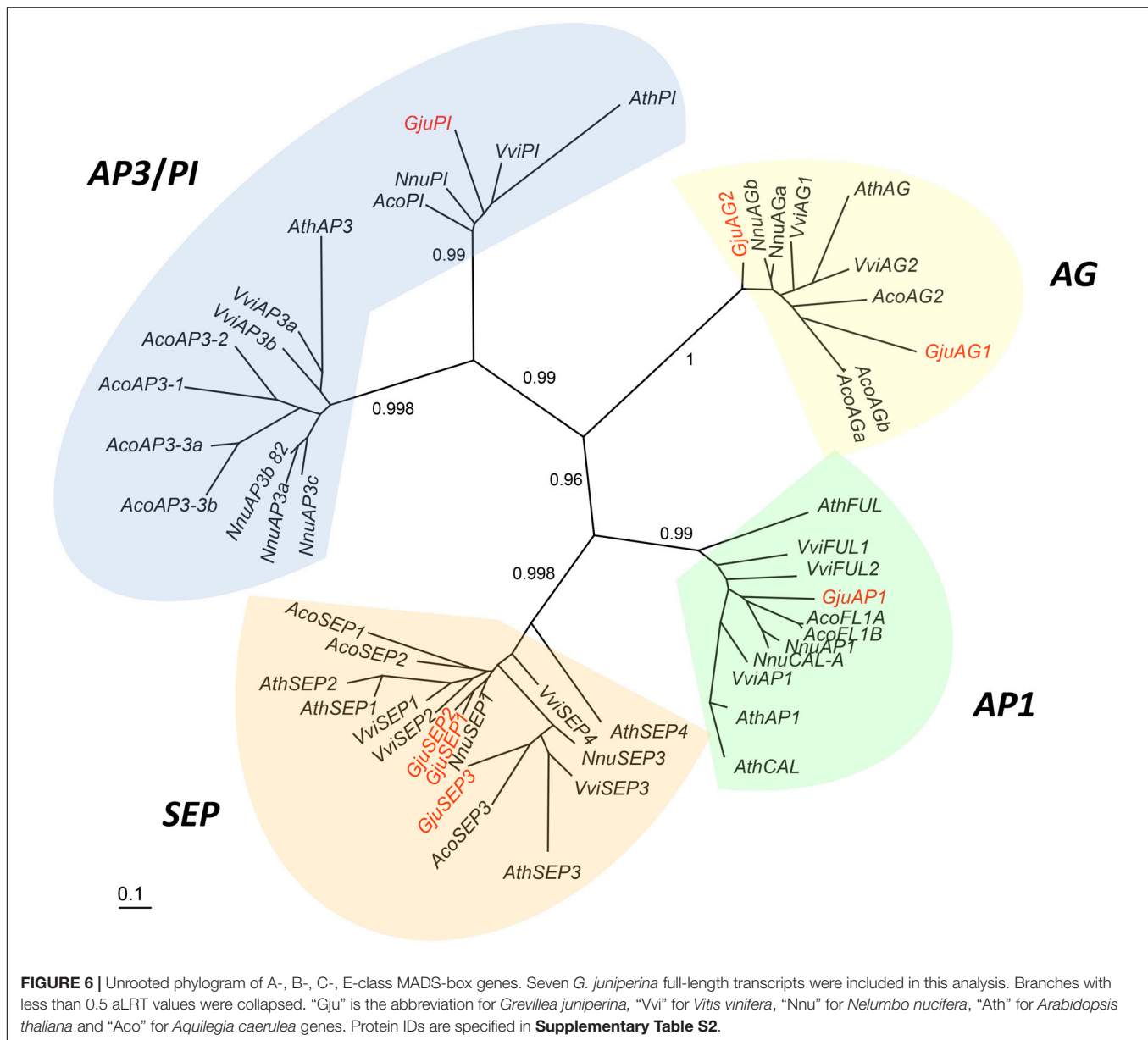
Homology With Other Eudicot Proteomes

Proteins deduced from the contigs of *G. juniperina* were compared to those in the proteomes of three fully sequenced species, *Arabidopsis thaliana*, *Nelumbo nucifera*, and *Vitis vinifera*, using to the algorithm available from the OrthoVenn site (Figure 4). As expected because of imperfect transcriptome *de*



novo assembly, the highest proportion of singletons was obtained from *G. juniperina* (58% of the number of proteins against 10–33% in the three other species). 20,787 clusters, i.e., orthologous groups were inferred, 14,362 of these included at least two species. 3,363 clusters included only sequences from *G. juniperina* while 1,132, 1,005, and 925 clusters were specific to *A. thaliana*, *V. vinifera*, and *N. nucifera*, respectively. This higher number of specific clusters could be due to various factors (distant homologs that were not identified, enrichment in specific floral

symmetry genes, assembly problems, etc). A GO enrichment test on these cluster members did not provide any clues as to a possible involvement in floral symmetry. 10,041 orthologous groups included sequences from all four species. Among the three pairwise comparisons involving the inferred *Grevillea* proteome, *Grevillea* and *Nelumbo*, both from the order Proteales, shared the highest number of total and pair-specific orthologous groups, which is consistent with their phylogenetic position. Orthologous groups including sequences of *Grevillea*, *Vitis*, and *Nelumbo* were



more numerous than those including sequences of *Arabidopsis*, *Vitis*, and *Nelumbo*.

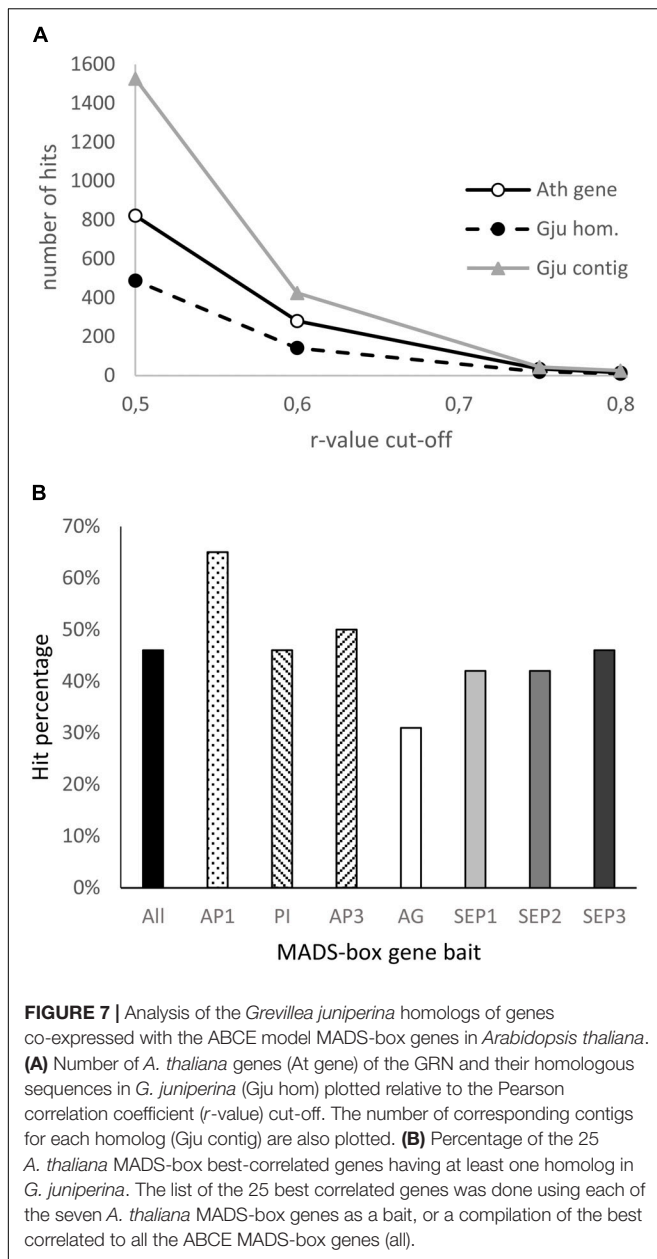
Mining the *Grevillea juniperina* Transcriptome for Homologs of Genes Involved in Floral Development

We focused on two families of transcription factors involved in growth and development processes, particularly in flowers. TCP genes have been reported to be involved in various cell proliferation and growth processes that control organ development and shape, including floral symmetry (Huang and Irish, 2015; Nicolas and Cubas, 2016). MADS-box genes of A-, B-, C-, and E- classes are involved in floral organ identity and also interact with patterning and growth genes (Sablowski, 2015).

In addition, we searched for homologs of genes co-expressed with the ABCE MADS-box genes as candidates for a conserved floral gene regulatory network module between *A. thaliana* and *G. juniperina*.

Phylogeny of TCP Homologous Genes

Twenty-four contigs with homology to TCP genes were identified in our *G. juniperina* transcriptome. Assembled fragment lengths ranged from 245 to 4,726 nucleotides; 10 fragments were predicted to contain complete ORFs (with identifiable start and stop codons) with putative ORF lengths ranging from 211 to 573 amino acids. All contigs but three had the characteristic basic helix-loop-helix TCP domain. Of the three contigs where the TCP domain could not be found, two had 76.4% sequence identity with R domains that were identified by blastx searches



as homologous to TCP2. One contig lacked any recognizable domain because it corresponded to the region downstream of the TCP domain, but was found to be homologous by blastx searches to TCP8.

The phylogenetic analysis of 192 aligned nucleotide positions, mainly comprising the TCP domain, from the 21 *G. juniperina* (excluding the three contigs without TCP domain) and 24 *A. thaliana* TCP genes showed that both class I and class II TCP were recovered with high support (Figure 5). Class I genes were over-represented in *G. juniperina*, with at least 14 copies compared to *A. thaliana* (13 copies), unlike class II genes with at least seven copies compared to *A. thaliana* (11 copies). One CYC-like copy (homologous to the CYC-like genes TCP18/TCP12/TCP1 in *A. thaliana*) was found in the

transcriptome of *G. juniperina*, corresponding to the previously characterized *ProtCYC1* (Citerne et al., 2017).

Phylogeny of ABCE MADS-Box Genes

Sixteen contigs annotated as one of the *Arabidopsis thaliana* ABCE MADS-box AGI gene identifiers were extracted from the *G. juniperina* transcriptome after annotation. Seven fragments were predicted to contain full length ORFs (with identifiable start and stop codons) with putative lengths ranging from 209 to 247 amino acids. All full-length contigs had the characteristic MIKC domains. Among the nine incomplete contigs, four lacked the N-terminal coding sequence and five lacked the C-terminal region. Among these incomplete sequences, one, three, and five contigs aligned partially with *GjuAPI*, *GjuSEP*, and *GjuPI*, respectively, suggesting other alleles or paralogs.

A phylogenetic tree including full length *G. juniperina* coding sequences and A-, B-, C-, and E-class MADS-box CDS from *Vitis vinifera*, *Nelumbo nucifera*, *Arabidopsis thaliana*, and *Aquilegia caerulea* was reconstructed using PhyML (Figure 6). We retrieved the A-, B-, C-, and E-class gene lineages with good support, and at least one *G. juniperina* sequence falls in each of these clades. Two AG-like *G. juniperina* sequences were found in the C lineage, one of which is closely related to the *N. nucifera* sequences. Three *G. juniperina* sequences were found in the E lineage. One of these fell in the SEP3 clade; the other two grouped together and might correspond to closely related paralogs. Surprisingly in the B-class, no *G. juniperina* sequence was found in the AP3 lineage, while one sequence was found in the PI one, closely related to the two other basal eudicot sequences.

Homologs of Genes Co-expressed With ABCE MADS-Box Genes

To investigate the extent to which the gene network involved in floral organ development is conserved between *Arabidopsis* and *Grevillea*, we searched for genes with expression patterns that are correlated with that of ABCE MADS-box genes in *Arabidopsis thaliana* using the Expression Angler Website, and then searched for their homologs in the annotated *G. juniperina* transcriptome. In *Arabidopsis*, 822 and 34 genes were found with a Pearson correlation coefficient cut-off (*r*-value) of, respectively, 0.5 and 0.75 (Supplementary Table S3). Analysis of the annotated *G. juniperina* transcriptome indicates global conservation of the gene network between the two species (Figure 7A) although with some variations (Table 1). Homologs of ABCE genes and of their highest correlated genes were found in the *G. juniperina* transcriptome, suggesting they are conserved between the two species (Figure 7A and Table 1). Indeed, 10 out of the 14 *Arabidopsis* genes with an *r*-value of 0.8 have at least one homologous contig in the *G. juniperina* transcriptome.

Analyses of the 25 best correlated genes to each of the seven *A. thaliana* ABCE MADS-box genes showed an overrepresentation of homologs of the AP1-correlated genes and an underrepresentation of the AG-correlated genes in the *G. juniperina* annotated transcriptome. The analysis confirmed that no homologs of AP3 were found in the *G. juniperina*

transcriptome (Table 1). Nevertheless, the percentage of *G. juniperina* homologs of AP3-correlated genes was similar to the percentage of the homologs of PI-correlated genes, or to the percentage computed over all the best ABCE MADS-box correlated genes (Figure 7B).

Gene Expression During Floral Development

Homologs of genes that play a role in floral meristem termination, organ identity/development and boundaries between organs in *Arabidopsis thaliana* were identified in the *G. juniperina* transcriptome annotated against the *Arabidopsis* proteome. In particular, genes expressed in carpel and/or ovule development in *A. thaliana* (Arnaud and Pautot, 2014;

Maugarny et al., 2016) and in other distantly related taxa (Fourquin et al., 2005; Reymond et al., 2012) were targeted, since the gynoecium can easily be separated from perianth with adnate stamens tissues, and its growth pattern is a major component of floral symmetry. In addition, the two previously characterized *ProtCYC1* and *ProtCYC2* genes (Citerne et al., 2017) respectively, *GjuCYC1* and *GjuCYC2*, were included in the expression analysis.

The expression pattern of these genes was examined in dissected organs from 6 to 7 mm buds using qRT-PCR. *GjuCRC* and *GjuSTM* had a predominant expression in the gynoecium compared to the dorsal and ventral organs, while *GjuAP2*, *GjuBEL1*, *GjuCYC2* and *GjuPI2* were predominantly expressed in the dorsal and ventral organs. *GjuSEP1* and *GjuSEP3* were significantly induced in the ventral organs compared to the

TABLE 1 | List of genes co-expressed with the ABCE MADS-box genes in *Arabidopsis thaliana* (r -value ≥ 0.75 , in bold the 14 AGI-ID with an r -value ≥ 0.8 ; AGI ID, Arabidopsis Genome Initiative, gene identifier).

AGI ID	Annotation	Gju. Contig number
With at least one homologous sequence in the <i>G. juniperina</i> transcriptome.		
At1g24260	AGL9_SEP3__K-box region and MADS-box transcription factor family protein	2
At3g02310	AGL4_SEP2__K-box region and MADS-box transcription factor family protein	4
At4g18960	AG__K-box region and MADS-box transcription factor family protein	1
At5g20240	PI__K-box region and MADS-box transcription factor family protein	6
At1g69120	AGL7_AP1_AtAP1__K-box region and MADS-box transcription factor family protein	1
At1g78960	ATLUP2_LUP2__lupeol synthase 2	6
At1g59640	BPE_BPEp_BPEub_ZCW32__BIG PETAL P	2
At1g02190	Fattyacid hydroxylase superfamily	1
At3g03450	RGL2__RGA-like 2	1
At1g05490	chr31__chromatin remodeling 31	1
<i>At2g38110</i>	ATGPAT6_GPAT6__glycerol-3-phosphate acyltransferase 6	4
<i>At5g02190</i>	ATASP38_EMB24_PCS1__Eukaryotic aspartyl protease family protein	1
<i>At3g01980</i>	NAD(P)-binding Rossmann-fold superfamily protein	1
<i>At5g59120</i>	ATSBT4.13_SBT4.13__subtilase 4.13	1
<i>At5g35670</i>	iqd33__IQ-domain 33	1
<i>At5g33370</i>	GDLS-like Lipase/Acylhydrolase superfamily protein	5
<i>At2g20870</i>	Cell wall protein precursor, putative	1
<i>At5g22430</i>	Pollen Ole e 1 allergen and extensin family protein	3
<i>At4g09960</i>	AGL11_STK__K-box region and MADS-box transcription factor family protein	1
With no homologous sequence in the <i>G. juniperina</i> transcriptome.		
At3g54340	AP3_ATAP3__K-box region and MADS-box transcription factor family protein	
At5g15800	AGL2_SEP1__K-box region and MADS-box transcription factor family protein	
At4g01080	TBL26__TRICHOME BIREFRINGENCE-LIKE 26	
At3g10570	CYP77A6__cytochrome P450, family 77, subfamily A, polypeptide 6	
<i>At1g16705</i>	p300/CBP acetyltransferase-related protein-related	
<i>At3g04960</i>	Domain of unknown function (DUF3444)	
<i>At1g66350</i>	RGL_RGL1__RGA-like 1	
<i>At4g21590</i>	ENDO3__endonuclease 3	
<i>At1g27360</i>	SPL11__squamosa promoter-like 11	
<i>At4g14695</i>	Uncharacterised protein family (UPF0041)	
<i>At3g28500</i>	60S acidic ribosomal protein family	
<i>At5g48880</i>	KAT5_PKT1_PKT2__peroxisomal 3-keto-acyl-CoA thiolase 2	
<i>At1g35170</i>	TRAM, LAG1, and CLN8 (TLC) lipid-sensing domain containing protein	
<i>At2g42830</i>	AGL5_SHP2__K-box region and MADS-box transcription factor family protein	
<i>At1g72260</i>	THI2.1_THI2.1.1__thionin 2.1	

gynoecium. Finally, *GjuCYC1* was predominantly expressed in the ventral organs (Figure 8).

DISCUSSION

By describing specific developmental features of flowers within Proteaceae, we found that the adnation of stamens to tepals takes place at early developmental stages, and that the establishment of bilateral symmetry coincides with asymmetrical growth of the single carpel. The transcriptome data obtained in parallel provides a wide array of genes expressed during floral development, opening the way for future studies to look at the expression of selected candidates within specific organs and/or at specific developmental stages, as discussed below.

Floral Development

In angiosperms, the initiation of floral organs typically follows a highly conserved pattern, with a centripetal initiation of the organs from the outermost to the innermost (generally the gynoecium) (Endress, 2006). However, there is no general rule as to when symmetry is established during development (Endress, 1999). In species with zygomorphic flowers, bilateral symmetry can be visible very early on during development; alternatively, floral buds can remain actinomorphic until late stages of development. Furthermore, species with actinomorphic flowers at anthesis may undergo transitory bilateral symmetry of the floral bud during development (reviewed in Reyes et al., 2016). At the gynoecium level, the type of symmetry is conditioned by the number of carpels and their closure pattern, which can be either ascidate or plicate (Endress, 2015; Sokoloff et al., 2017). In species with a single plicate carpel presenting a marked cleft during development, as it is the case in Proteaceae (Douglas and Tucker, 1996b), the gynoecium becomes zygomorphic as soon as the cleft begins to form. In most subfamilies of Proteaceae, the cleft is typically oriented along the dorso-ventral axis, facing the adaxial tepal. In Grevilloideae, however, the situation is more complex, as described in detail by Douglas and Tucker (1996b) (see also Johnson and Briggs, 1975), who suggested that the variation in carpel orientation could be related to the space left after stamen initiation. Our observations suggest asymmetrical growth of the carpel primordium from the very early stages, resulting in early zygomorphy of the gynoecium. The orientation of flowers within the flower pair is also variable (Douglas and Tucker, 1996a), making the interpretation of axes of symmetry even more complex. Such variation in carpel and flower orientation has been suggested to be related to the inflorescence structure and degree of compaction that are both variable across the family (Johnson and Briggs, 1975). In Embothrieae, the tribe to which our four study species belong, zygomorphy is widespread and has been inferred as the ancestral condition (Citerne et al., 2017). The different degrees of floral bilateral symmetry at anthesis in the species examined here could be related to differences in inflorescence morphology. How the shape and compaction of the inflorescence may condition flower symmetry remains to be explored. The

fact that *ProtCYC* genes are expressed (although faintly in the case of *ProtCYC1*) in the gynoecium of *G. juniperina* (Citerne et al., 2017 and this study), and asymmetrically in the tepals and/or adnate stamens (*ProtCYC1* only), suggests that these genes could play a role in the genetic control of floral zygomorphy in Proteaceae, as already suggested by Citerne et al. (2017).

The position of stamens, opposite the tepals, questions the identity of the perianth. In whorled flowers, the stamens usually alternate with adjacent perianth parts when in equal number. In the order Ericales, obhaptostemony (petal-opposed stamens) is believed to be derived from diplostemony (Schönenberger et al., 2005), with the loss of a stamen whorl. Detailed expression studies of the homologs of floral organ identity genes and their correlated genes found in the *G. juniperina* transcriptome may provide clues as to the homology of the perianth. Furthermore, the transcriptomic data may also help to investigate the genetic control of organ fusion.

A Transcriptomic Tool to Investigate the Genetic Bases of Floral Development

Transcriptomic data from Proteaceae species are scarce and have mostly been obtained from leaves for specific purposes: in *Leucadendron* to derive phylogenetic markers (Tonnabel et al., 2014), in *Protea repens* to study population differentiation in relation to local adaptation (Akman et al., 2016), in *Banksia hookeriana* to derive SSR markers (Lim et al., 2017) and in *Gevuina avellana* to study heteroblastic development under different light conditions (Ostria-Gallardo et al., 2016). In *Macadamia integrifolia*, a reference transcriptome was built from flowers, shoots and leaves, in parallel to a draft genome (Nock et al., 2016). Our *Grevillea juniperina* transcriptome was built from an equal mix of RNAs from young leaves and flower buds at three developmental stages, all posterior to organ inception. The general metrics of this transcriptome such as N50, mean length of contigs, percentage of annotated contigs, all fall within the range of values obtained in other Proteaceae species (e.g., Akman et al., 2016). The sequencing effort and mix of tissues were designed to cover a large panel of expressed genes including transcription factors, with a higher proportion of floral samples in the library used for RNA-seq to obtain a good coverage of transcripts expressed during floral development. We probably missed some rare floral transcripts, such as the *ProtCYC2* gene that we previously found to be expressed in floral buds of the same species of *Grevillea* (Citerne et al., 2017), and whose expression was confirmed here. Equally, the absence of a homolog of the B-class gene *AP3* is intriguing, because we were able to find at least one homologous contig for all the other A-, B-, C-, and E-class genes. The fact that we found *G. juniperina* homologs of *A. thaliana* genes with correlated expression with *AP3* would rather be in favor of a technical bias. Gene expression analysis showed expected patterns for the homologs of the identity genes tested, including *GjuAP2*, a homolog of the *A. thaliana* *AP2* gene, which is one of the two A-class genes and the only ABCE model gene that is not from the MADS-box gene

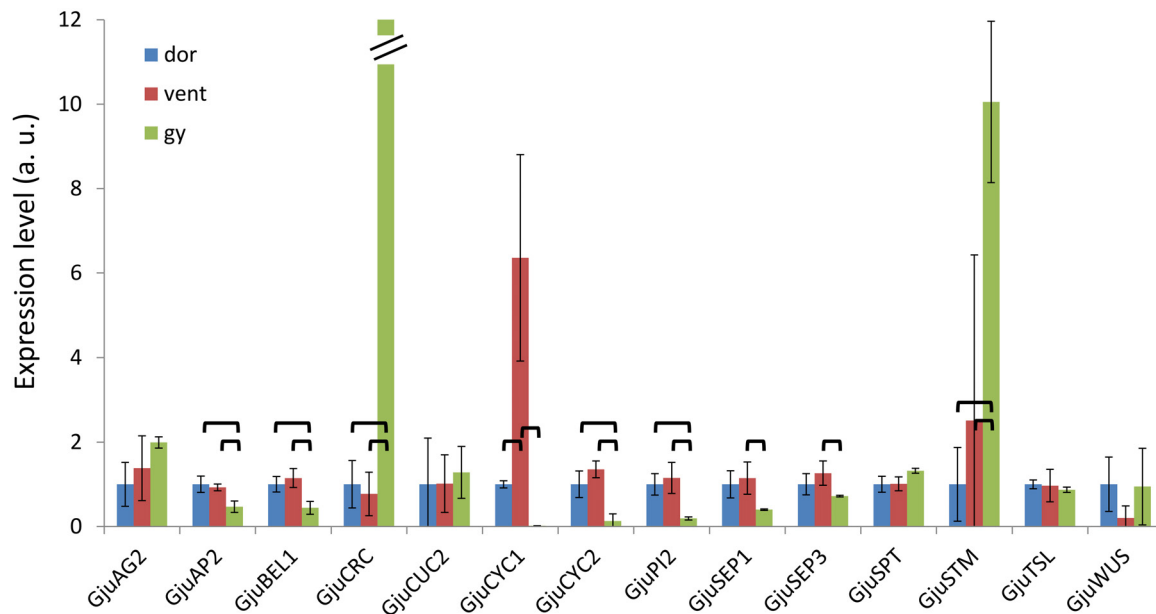


FIGURE 8 | Expression patterns of 14 floral genes assessed by qRT-PCR in floral dissections of *Grevillea juniperina*. Organs were dissected from 6–7 mm buds; dor, dorsal tepals and adnate stamens; vent, ventral tepals and adnate stamens; gy, gynoecium. For a better display, the normalized expression levels of each gene in this figure are shown relative to the dor sample. In this figure, the expression level of *GjuCRC* in the gynoecium is 844. The error bars are calculated from three biological replicates. The statistically significant pairwise comparisons are indicated with brackets.

family. TCP genes are also well represented in the transcriptome, with 14 and seven homologs belonging to class I and class II subfamilies, respectively, which is comparable to the size of the family in many eudicots. One sequence matched the previously characterized *ProtCYC1* gene, which is homologous to the *Arabidopsis* CYC-like genes (*TCP18/TCP12/TCP1*). This gene is the only one exhibiting an asymmetric expression pattern in the floral dissection that could be related to the asymmetry of the perianth and stamens, as already observed by Citerne et al. (2017) in this species and another zygomorphic *Grevillea* species.

ABCE MADS-box genes are master regulators of the floral organ development Gene Regulatory Network (GRN) that are subjected to strong functional constraints, in contrast with the great flexibility of organ morphogenesis (Davila-Velderrain et al., 2013; Becker and Ehlers, 2016). Genes co-expressed with these regulatory genes in *Arabidopsis* could serve as baits to identify conserved modules of connected genes in other species and analyze their evolution and expression patterns. A large proportion of genes with expression patterns that are highly correlated with floral organ identity gene expression in *Arabidopsis thaliana* have homologs in the *G. juniperina* transcriptome (10 out of 14). Available functional studies in *A. thaliana* indicate that these genes code for transcription factors and/or hormonal signaling pathway proteins, and are important regulators of flower organ development, sexual reproduction, and seed fertility (Ge et al., 2005; Szecsi et al., 2006; Hou et al., 2008; Varaud et al., 2011; Li et al., 2012; Hong et al., 2017). A less stringent correlation threshold gives a lesser proportion of homologous genes in the *G. juniperina* transcriptome. This may

suggest that beyond a core set of genes whose expression patterns are strongly parallel to that of organ identity genes, other genes involved in flower development have diverged in the two species. It is worth noting the relative overrepresentation in *G. juniperina* of homologs of *API*-correlated genes. It remains to be seen whether this finding may be relevant for assessing the homology of undifferentiated perianth organs with sepals or petals. Among the genes expressed during carpel and ovule development in *Arabidopsis*, only *GjuCRC* and *GjuSTM* exhibit a predominant expression in the gynoecium of *G. juniperina*, while *GjuAG2* was also found to be expressed in perianth and stamen tissues (both ventral and dorsal organs). Detailed expression analysis of these genes at the tissue level (using *in situ* hybridization for example) may provide insight into the processes involved in the monosymmetric growth of the gynoecium during floral ontogeny.

Floral evo-devo questions that concern Proteaceae are many, among which the homology of the undifferentiated perianth to a calyx or a corolla, the adnation of stamens, or the chronology and developmental processes of the acquisition of zygomorphy. Such issues could be addressed by combining developmental and molecular approaches, for example using RNA-seq at different floral developmental stages in a panel of well-chosen species. The *G. juniperina* reference transcriptome will be a valuable resource for all these future studies. It will be a reference for identifying differentially expressed genes related to the establishment of symmetry at appropriate developmental stages in different floral organs, and for investigating the gene network underlying tepal development, providing insights into the homology of the colored perianth organs.

Choosing appropriate developmental stages may be facilitated by High Resolution X-Ray Computed Tomography giving access to early developmental stages without dissection, which is especially valuable for species with young compact and tough inflorescences.

AUTHOR CONTRIBUTIONS

SN and CD designed the project. HS, FS, YS, JS, and SN performed the developmental analysis. ED and JY generated RNA-seq data and transcriptome assembly. JJ, NCS, YD, HC, CD, and FS annotated and mined transcriptomic data. ED and MLG performed the expression analysis. All authors contributed to the manuscript.

FUNDING

This work was funded by a PRES UNIVERSUD grant for the SPAM project, a 2012 grant from IFR 87 ‘La Plante et son Environnement’ and by the Agence Nationale de la Recherche

(ANR-07-BLAN-0112 grant). HC and FS were funded by ANR and PRES grants, respectively.

ACKNOWLEDGMENTS

We gratefully acknowledge Peter Weston and Peter Olde for providing some of the plant material, Véronique Brunaud for her work on the first version of the transcriptome assembly, Florian Jabbour and Dmitry Sokoloff for helpful discussions on floral developmental, and Véronique Normand for her technical assistance in the lab.

The GQE-Le Moulon and Institute of Plant Sciences Paris Saclay benefit from the support of the LabExSaclay Plant Sciences-SPS (ANR-10-LABX-0040-SPS).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.00018/full#supplementary-material>

REFERENCES

- Akman, M., Carlson, J. E., Holsinger, K. E., and Latimer, A. M. (2016). Transcriptome sequencing reveals population differentiation in gene expression linked to functional traits and environmental gradients in the South African shrub *Protea repens*. *New Phytol.* 210, 295–309. doi: 10.1111/nph.13761
- Arnaud, N., and Pautot, V. (2014). Ring the BELL and tie the KNOX: roles for TALES in gynoecium development. *Front. Plant Sci.* 5:93. doi: 10.3389/fpls.2014.00093
- Bartlett, M. E., Williams, S. K., Taylor, Z., DeBlasio, S., Goldshmidt, A., Hall, D. H., et al. (2015). The maize P1/GLO ortholog Zmm16/sterile tassel silky ear1 interacts with the zygomorphy and sex determination pathways in flower development. *Plant Cell* 27, 3081–3098. doi: 10.1105/tpc.15.00679
- Becker, A., and Ehlers, K. (2016). Arabidopsis flower development—of protein complexes, targets, and transport. *Protoplasma* 253, 219–230. doi: 10.1007/s00709-015-0812-7
- Broholm, S. K., Tähtiharju, S., Laitinen, R. A. E., Albert, V. A., Heeri, T. H., and Elomaa, P. (2008). A TCP domain transcription factor controls flower type specification along the radial axis of the *Gerbera* (Asteraceae) inflorescence. *Proc. Nat. Acad. Sci. U.S.A.* 105, 9117–9122. doi: 10.1073/pnas.0801359105
- Busch, A., and Zachgo, S. (2007). Control of corolla monosymmetry in the Brassicaceae *Iberis amara*. *Proc. Nat. Acad. Sci. U.S.A.* 104, 16714–16719. doi: 10.1073/pnas.0705338104
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421
- Causier, B., Schwarz-Sommer, Z., and Davies, B. (2010). Floral organ identity: 20 years of ABCs. *Sem. Cell Dev. Biol.* 21, 73–79. doi: 10.1016/j.semcdb.2009.10.005
- Citerne, H. L., Reyes, E., Le Guilloux, M., Delannoy, E., Simonnet, F., Sauquet, H., et al. (2017). Characterization of CYCLOIDEA-like genes in Proteaceae, a basal eudicot family with multiple shifts in floral symmetry. *Ann. Bot.* 119, 367–378. doi: 10.1111/nph.13761
- Clark, J. I., and Coen, E. S. (2002). The cycloidea gene can respond to a common dorsoventral prepattern in Antirrhinum. *Plant J.* 30, 639–648. doi: 10.1046/j.1365-3113X.2002.01310.x
- Coen, E., and Meyerowitz, E. M. (1991). The war of the whorls: genetic interactions controlling flower development. *Nature* 353, 31–37. doi: 10.1038/353031a0
- Cubas, P., Lauter, N., Doebley, J., and Coen, E. (1999). The TCP domain: a motif found in proteins regulating plant growth and development. *Plant J.* 18, 215–222. doi: 10.1046/j.1365-3113X.1999.00444.x
- Davila-Velderrain, J., Servin-Marquez, A., and Alvarez-Buylla, E. R. (2013). Molecular evolution constraints in the floral organ specification gene regulatory network module across 18 angiosperm genomes. *Mol. Biol. Evol.* 31, 560–573. doi: 10.1093/molbev/mst223
- Douglas, A. W., and Tucker, S. C. (1996a). Inflorescence ontogeny and floral organogenesis in Grevilleoideae (Proteaceae), with emphasis on the nature of the flower pairs. *Int. J. Plant Sci.* 157, 341–372. doi: 10.1086/297354
- Douglas, A. W., and Tucker, S. C. (1996b). The developmental basis of diverse carpel orientations in Grevilleoideae. *Int. J. Plant Sci.* 157, 373–397. doi: 10.1086/297355
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Endress, P. K. (1999). Symmetry in flowers: diversity and evolution. *Int. J. Plant Sci.* 160, S3–S23. doi: 10.1086/314211
- Endress, P. K. (2006). Angiosperm floral evolution: morphological developmental framework. *Adv. Bot. Res.* 44, 1–61. doi: 10.1016/S0065-2296(06)44001-5
- Endress, P. K. (2015). Patterns of angiosperm development before carpel: sealing across living angiosperms: diversity, and morphological and systematic aspects. *Bot. J. Lin. Soc.* 178, 556–591. doi: 10.1111/boj.12294
- Feng, X., Zhao, Z., Tian, Z., Xu, S., Luo, Y., Cai, Z., et al. (2006). Control of petal shape and floral zygomorphy in *Lotus* gynoecium. *Proc. Natl. Acad. Sci. U.S.A.* 103, 49670–44975. doi: 10.1073/pnas.0600681103
- Fourquin, C., Vinauger-Douard, M., Folgiani, B., Dumas, C., and Scutt, C. P. (2005). Evidence that CRABS CLAW and TOUSLED have conserved their roles in carpel development since the ancestor of the extant angiosperms. *Proc. Natl. Acad. Sci. U.S.A.* 102, 4649–4654. doi: 10.1073/pnas.0409577102
- Ge, X., Dietrich, C., Matsuno, M., Li, G., Berg, H., and Xia, Y. (2005). An *Arabidopsis* aspartic protease functions as an anti-cell-death component in reproduction and embryogenesis. *EMBO Rep.* 6, 282–288. doi: 10.1038/sj.embor.7400357
- 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
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/nprot.2013.084
- Hileman, L. C. (2014). Trends in flower symmetry evolution revealed through phylogenetic and developmental genetic advances. *Philos. Trans. R. Soc. B* 369:20130348. doi: 10.1098/rstb.2013.0348

- Hong, L., Brown, J., Segerson, N. A., Rose, J. K., and Roeder, A. H. (2017). CUTIN SYNTHASE 2 maintains progressively developing cuticular ridges in *Arabidopsis* sepals. *Mol. Plant* 10, 560–574. doi: 10.1016/j.molp.2017.01.002
- Hou, X., Hu, W. W., Shen, L., Lee, L. Y., Tao, Z., Han, J. H., et al. (2008). Global identification of DELLA target genes during *Arabidopsis* flower development. *Plant Physiol.* 147, 1126–1142. doi: 10.1104/pp.108.121301
- Huang, T., and Irish, V. (2015). Temporal control of plant organ growth by TCP transcription factors. *Curr. Biol.* 25, 1765–1770. doi: 10.1016/j.cub.2015.05.024
- Johnson, L. A., and Briggs, B. G. (1975). On the Proteaceae – the evolution and classification of a southern family. *Bot. J. Lin. Soc.* 70, 83–182. doi: 10.1111/j.1095-8339.1975.tb01644.x
- Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031
- Lefort, V., Longueville, J. E., and Gascuel, O. (2017). SMS: smart Model Selection in PhyML. *Mol. Biol. Evol.* 34, 2422–2424. doi: 10.1093/molbev/msx149
- Li, X. C., Zhu, J., Yang, J., Zhang, G. R., Xing, W. F., Zhang, S., et al. (2012). Glycerol-3-phosphate acyltransferase 6 (GPAT6) is important for tapetum development in *Arabidopsis* and plays multiple roles in plant fertility. *Mol. Plant* 5, 131–142. doi: 10.1093/mp/ssr057
- Lim, L. S., D'Agui, H. M., Enright, N. J., and He, T. (2017). Characterization of leaf transcriptome in *Banksia hookeriana*. *Genomics Proteomics Bioinformatics* 15, 49–56. doi: 10.1016/j.gpb.2016.11.001
- Luo, D., Carpenter, R., Copsey, L., Vincent, C., Clark, J., and Coen, E. (1999). Control of organ asymmetry in flowers of *Antirrhinum*. *Cell* 99, 367–376. doi: 10.1016/S0092-8674(00)81523-8
- Luo, D., Carpenter, R., Vincent, C., Copsey, L., and Coen, E. (1996). Origin of floral asymmetry in *Antirrhinum*. *Nature* 383, 794–799. doi: 10.1038/383794a0
- Maugarny, A., Gonçalves, B., Arnaud, N., and Laufs, P. (2016). “CUC transcription factors: to the meristem and beyond,” in *Plant Transcription Factors. Evolutionary, Structural, and Functional Aspects*, ed. D. H. Gonzalez (Cambridge, MA: Academic Press), 229–247. doi: 10.1016/B978-0-12-800854-6.00015-4
- Mondragon-Palomino, M. (2013). Perspectives on MADS-box expression during orchid flower evolution and development. *Front. Plant Sci.* 4:377. doi: 10.3389/fpls.2013.00377
- Nakasugi, K., Crowhurst, R., Bally, J., and Waterhouse, P. (2014). Combining transcriptome assemblies from multiple *de novo* assemblers in the allotetraploid plant *Nicotiana benthamiana*. *PLoS One* 9:e91776. doi: 10.1371/journal.pone.0091776
- Nicolas, M., and Cubas, P. (2016). TCP factors: new kids on the signaling block. *Curr. Opin. Plant Biol.* 33, 33–41. doi: 10.1016/j.pbi.2016.05.006
- Nock, C. J., Baten, A., Barkla, B. J., Furtado, A., Henry, R. J., and King, G. J. (2016). Genome and transcriptome sequencing characterizes the gene space of *Macadamia integrifolia* (Proteaceae). *BMC Genomics* 17:937. doi: 10.1186/s12864-016-3272-3
- Ostria-Gallardo, E., Ranjan, A., Zumstein, K., Chitwood, D. H., Kumar, R., Townsley, B. T., et al. (2016). Transcriptomic analysis suggests a key role for SQUAMOSA PROMOTER BINDING PROTEIN LIKE, NAC and YUCCA genes in the heteroblastic development of the temperate rainforest tree *Gevuina avellana* (Proteaceae). *New Phytol.* 210, 694–708. doi: 10.1111/nph.13776
- Preston, J. C., and Hileman, L. C. (2012). Parallel evolution of TCP and B-class genes in Commelinaceae flower bilateral symmetry. *EvoDevo* 3:6. doi: 10.1186/2041-9139-3-6
- Reyes, E., Sauquet, H., and Nadot, S. (2016). Perianth symmetry changed at least 199 times in angiosperm evolution. *Taxon* 65, 945–964. doi: 10.12705/655.1
- Reymond, M. C., Brunoud, G., Chauvet, A., Martinez-Garcia, J. F., Martin-Magniette, M.-L., Monéger, F., et al. (2012). A light-regulated genetic module was recruited to carpel development in *Arabidopsis* following a structural change to SPATULA. *Plant Cell* 24, 2812–2825. doi: 10.1105/tpc.112.097915
- Roberts, W. R., and Roalson, E. H. (2017). Comparative transcriptome analyses of flower development in four species of *Achimenes* (Gesneriaceae). *BMC Genomics* 18:240. doi: 10.1186/s12864-017-3623-8
- Sablowski, R. (2015). Control of patterning, growth, and differentiation by floral organ identity genes. *J. Exp. Bot.* 66, 1065–1073. doi: 10.1093/jxb/eru514
- Sauquet, H., von Balthazar, M., Magallon, S., Doyle, J. A., Endress, P. K., Bailes, E. J., et al. (2017). The ancestral flower of angiosperms and its early diversification. *Nat. Commun.* 8:16047. doi: 10.1038/ncomms16047
- Schönenberger, J., Anderberg, A. A., and Sytsma, K. J. (2005). Molecular phylogenetics and patterns of floral evolution in the Ericales. *Int. J. Plant Sci.* 166, 265–288. doi: 10.1086/427198
- Schulz, M. H., Zerbino, D. R., Vingron, M., and Birney, E. (2012). Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 28, 1086–1092. doi: 10.1093/bioinformatics/bts094
- Smart, M., and Roden, L. C. (2010). A small-scale RNA isolation protocol useful for high-throughput extractions from recalcitrant plants. *S. Afr. J. Bot.* 76, 375–379. doi: 10.1016/j.sajb.2010.01.002
- Sokoloff, D. D., Nuraliev, M. S., Oskolski, A. A., and Remizowa, M. V. (2017). Gynoecium evolution in angiosperms: monomery, pseudomonomery, and mixomery. *Moscow Univ. Biol. Sci. Bulletin* 72, 97–108. doi: 10.3732/ajb.1600269
- Spencer, V., and Kim, M. (2018). Re“CYC”ling molecular regulators in the evolution and development of flower symmetry. *Semin. Cell Dev. Biol.* 79, 16–26. doi: 10.1016/j.semdb.2017.08.052
- Staedler, Y. M., Masson, D., and Schönenberger, J. (2013). Plant tissues in 3D via X-ray tomography: simple contrasting methods allow high resolution imaging. *PLoS One* 8:e75295. doi: 10.1371/journal.pone.0075295
- Szecs, J., Joly, C., Bordji, K., Varaud, E., Cock, J. M., Dumas, C., et al. (2006). BIGPETA1, a bHLH transcription factor is involved in the control of *Arabidopsis* petal size. *EMBO J.* 25, 3912–3920. doi: 10.1038/sj.emboj.7601270
- Theißen, G., Kim, J. T., and Saedler, H. (1996). Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes. *J. Mol. Evol.* 43, 484–516. doi: 10.1007/BF02337521
- Tonnabel, J., Olivieri, I., Mignot, A., Rebelo, A., Justy, F., Santoni, S., et al. (2014). Developing nuclear DNA phylogenetic markers in the angiosperm genus *Leucadendron* (Proteaceae): a next-generation sequencing transcriptomic approach. *Mol. Phylog. Evol.* 70, 37–46. doi: 10.1016/j.ympev.2013.07.027
- Toufighi, K., Brady, S. M., Austin, R., Ly, E., and Provart, N. J. (2005). The Botany Array Resource: e-Northern, expression angling, and promoter analyses. *Plant J.* 43, 153–163. doi: 10.1111/j.1365-313X.2005.02437.x
- Tucker, S. C. (1999). Evolutionary lability of symmetry in early floral development. *Int. J. Plant Sci.* 160(6 Suppl.), S25–S39. doi: 10.1086/314212
- Varaud, E., Brioudes, F., Szecs, J., Leroux, J., Brown, S., Perrot-Rechenmann, C., et al. (2011). AUXIN RESPONSE FACTOR8 regulates *Arabidopsis* petal growth by interacting with the bHLH transcription factor BIGPETA1. *Plant Cell* 23, 973–983. doi: 10.1105/tpc.110.081653
- Wang, Y., Coleman-Derr, D., Chen, G., and Gu, Y. Q. (2015). OrthoVenn : a web server for genome wide comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res.* 43, W78–W84. doi: 10.1093/nar/gkv487
- Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., et al. (2018). BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* 35, 543–548. doi: 10.1093/molbev/msx319
- Weston, P. H. (2007). “Proteaceae,” in *The Families and Genera of Vascular Plants*, Vol. 9, ed. K. Kubitski (Berlin: Springer), 364–404. doi: 10.1007/978-3-540-32219-1_42
- Weston, P. H., and Barker, N. P. (2006). A new suprageneric classification of the Proteaceae, with an annotated checklist of genera. *Telopea* 11, 314–344. doi: 10.7751/teleopea20065733
- Yuan, Z., Gao, S., Xue, D.-W., Luo, D., Li, L.-T., Ding, S.-Y., et al. (2009). RETARDED PALEA1 controls palea development and floral zygomorphy in rice. *Plant Physiol.* 149, 235–244. doi: 10.1104/pp.108.128231

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 Damerval, Citerne, Conde e Silva, Deveaux, Delannoy, Joets, Simonnet, Staedler, Schönenberger, Yansouni, Le Guilloux, Sauquet and Nadot. 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.



Novel Traits, Flower Symmetry, and Transcriptional Autoregulation: New Hypotheses From Bioinformatic and Experimental Data

Aniket Sengupta and Lena C. Hileman*

The Hileman Lab, Department of Ecology and Evolutionary Biology, The University of Kansas, Lawrence, KS, United States

OPEN ACCESS

Edited by:

Verónica S. Di Stilio,
University of Washington,
United States

Reviewed by:

Lydia Gramzow,
Friedrich-Schiller-Universität Jena,
Germany
Ana Maria Rocha De Almeida,
California State University, East Bay,
United States

*Correspondence:

Lena C. Hileman
lhileman@ku.edu

Specialty section:

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

Received: 11 May 2018

Accepted: 05 October 2018

Published: 26 October 2018

Citation:

Sengupta A and Hileman LC (2018)
Novel Traits, Flower Symmetry,
and Transcriptional Autoregulation:
New Hypotheses From Bioinformatic
and Experimental Data.
Front. Plant Sci. 9:1561.
doi: 10.3389/fpls.2018.01561

A common feature in developmental networks is the autoregulation of transcription factors which, in turn, positively or negatively regulate additional genes critical for developmental patterning. When a transcription factor regulates its own expression by binding to *cis*-regulatory sites in its gene, the regulation is direct transcriptional autoregulation (DTA). Indirect transcriptional autoregulation (ITA) involves regulation by proteins expressed downstream of the target transcription factor. We review evidence for a hypothesized role of DTA in the evolution and development of novel flowering plant phenotypes. We additionally provide new bioinformatic and experimental analyses that support a role for transcriptional autoregulation in the evolution of flower symmetry. We find that 5' upstream non-coding regions are significantly enriched for predicted autoregulatory sites in Lamiales *CYCLOIDEA* genes—an upstream regulator of flower monosymmetry. This suggests a possible correlation between autoregulation of *CYCLOIDEA* and the origin of monosymmetric flowers near the base of Lamiales, a pattern that may be correlated with independently derived monosymmetry across eudicot lineages. We find additional evidence for transcriptional autoregulation in the flower symmetry program, and report that *Antirrhinum DRIF2* may undergo ITA. In light of existing data and new data presented here, we hypothesize how *cis*-acting autoregulatory sites originate, and find evidence that such sites (and DTA) can arise subsequent to the evolution of a novel phenotype.

Keywords: *CYCLOIDEA*, evolution, flower development, symmetry, transcriptional autoregulation

INTRODUCTION

A common feature in developmental networks is the autoregulation of transcription factors which, in turn, positively or negatively regulate additional genes critical for developmental patterning. A *trans*-acting protein is considered transcriptionally autoregulated when the protein itself, or downstream factors, modulate its expression. Transcriptional autoregulation can be either direct, or indirect. In direct transcriptional autoregulation (DTA), a protein binds to *cis*-regulatory sites in its gene and modulates expression. Indirect transcriptional autoregulation (ITA) involves regulation by proteins expressed downstream of the target transcription factor (**Figure 1**). Both DTA and ITA have the potential to enter run-away

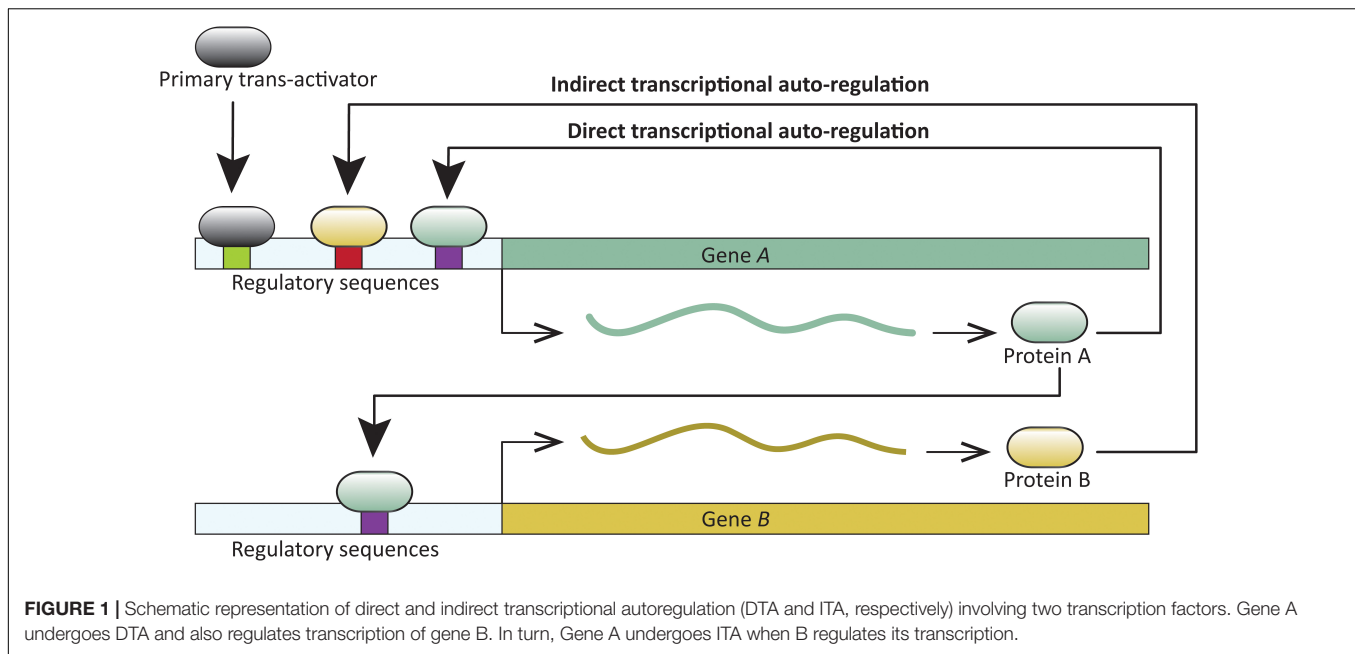


FIGURE 1 | Schematic representation of direct and indirect transcriptional autoregulation (DTA and ITA, respectively) involving two transcription factors. Gene A undergoes DTA and also regulates transcription of gene B. In turn, Gene A undergoes ITA when B regulates its transcription.

positive feedback processes. Expression of such genes is likely reduced or stabilized by additional regulatory factors. Transcription factor autoregulation is widespread. For example, at least 40% of transcription factors in *Escherichia coli* are autoregulated (Rosenfeld et al., 2002), and similar direct and indirect autoregulation has been reported across the tree of life—in viruses, prokaryotes, and eukaryotes (for example, Hochschild, 2002; Martínez-Antonio and Collado-Vides, 2003; Holloway et al., 2011; Tao et al., 2012; Gallo-Ebert et al., 2013; and reviewed in Bateman, 1998; Crews and Pearson, 2009), including those with complex development (for example, Cripps et al., 2004; Holloway et al., 2011; Ye et al., 2016). DTA has been demonstrated in processes as diverse, and crucial as the origin of certain cancers (Pasqualucci et al., 2003), and the onset of flowering (Tao et al., 2012).

The widespread occurrence of transcription factor autoregulation suggests a beneficial role in the function and evolution of genetic programs. Here, we provide a review of evidence for DTA in key flowering plant developmental programs. We provide new data supporting the hypothesis that DTA facilitated the evolution of flower monosymmetry in Lamiales. Together these data provide compelling evidence for the hypothesis that DTA plays a role in facilitating the evolution of novelty.

ADVANTAGES OF AUTOREGULATION

Several models suggest that autoregulation, especially DTA, can maintain a steady level of expression independent of other factors. If so, genes that are more likely to be autoregulated should be those that experience fleeting regulatory signals, or are positioned upstream in genetic regulatory networks with crucial developmental functions (Crews and Pearson, 2009;

Singh and Hespanha, 2009). For example, several transcription factors involved in antibiotic resistance are reported to be autoregulated, resistance being a crucial phenotype (Hoot et al., 2010; Hervay et al., 2011). Similarly, entering or exiting lytic and lysogenic stages is a key developmental decision in lambda bacteriophages, and this decision is partly controlled by the autoregulation of a transcription factor, CI (Hochschild, 2002). The prediction that transcription factors upstream in regulatory networks are more likely to undergo autoregulation has been tested in the model eukaryote yeast, *Saccharomyces cerevisiae*. In yeasts, where all possible transcription factor interactions have either been tested or predicted, master regulatory genes are significantly more likely to experience autoregulation than are other regulators (Odom et al., 2006). Similarly, five out of six master regulatory genes in human hepatocytes bind to their own promoters, i.e., undergo DTA (Odom et al., 2006).

How regulatory networks define stable phenotypes is an important question in evolution and development. Simulations of developmental network evolution suggest that autoregulated genes are more robust when faced with random mutations and environmental perturbation (Pinho et al., 2014). The model that DTA stabilizes expression by reducing system noise has been tested in the gene *hunchback* in *Drosophila melanogaster*. Models where the HOX transcription factor Hunchback binds to the *hunchback* promoter (i.e., *hunchback* undergoes DTA) predict less promoter binding-unbinding noise, making the system more robust (Holloway et al., 2011). Experimental work in *hunchback* mutants whose protein cannot bind to DNA (hence, cannot undergo DTA) supports this prediction (Holloway et al., 2011).

In addition to enhancing system robustness, autoregulation provides a mechanism for maintaining expression through key stages of development (reviewed below) that are potentially

critical for patterning phenotype. However, the developmental role of DTA has only been tested by mutational studies in a handful of cases. To determine the role of transcription factor DTA, the direct binding between the protein product of a gene and that gene's *cis*-regulatory DNA can be either intensified or weakened through direct DNA manipulation. For example, addition or deletion of *cis*-regulatory self-binding sites can be used to test for the specific developmental role of DTA within a given species (Espley et al., 2009; Tao et al., 2012; Gallo-Ebert et al., 2013). A complementary, but more difficult approach is to alter transcription factor peptide sequence by mutagenesis in order to modify affinity toward the self-binding sites, e.g., in the *hunchback* exemplified discussed above (Holloway et al., 2011). In some model systems, it is possible to repress activity of a transcription factor by overexpressing a dominant chimeric version of the peptide with a repressor domain added to the carboxy-terminus. The chimeric protein can repress the function of the native transcription factor by competitive inhibition (for example, Hiratsu et al., 2003; Koyama et al., 2010). Recent advances in CRISPR/Cas9 gene-editing

technologies (Ma et al., 2016) will certainly facilitate exploration of DTA function, at least in model species.

REVIEW OF DTA IN FLOWERING PLANT DEVELOPMENTAL EVOLUTION

Once an initial signal for activation of gene expression has been received, a transcription factor capable of DTA can contribute to swift developmental decisions. A clear example comes from work on the developmental transition to flowering (Figure 2A). Flowering time is a key life-history transition in plant development, intimately tied to environmental cues and aging in order to ensure reproductive success (reviewed in Ó'Maoláidigh et al., 2014). In *Arabidopsis thaliana*, the transition from vegetative to reproductive development is regulated in part by a MADS-box transcription factor, SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1). SOC1 undergoes DTA through the binding of SOC1 protein to four *cis*-regulatory *CARG-box* self-binding sites close to the SOC1

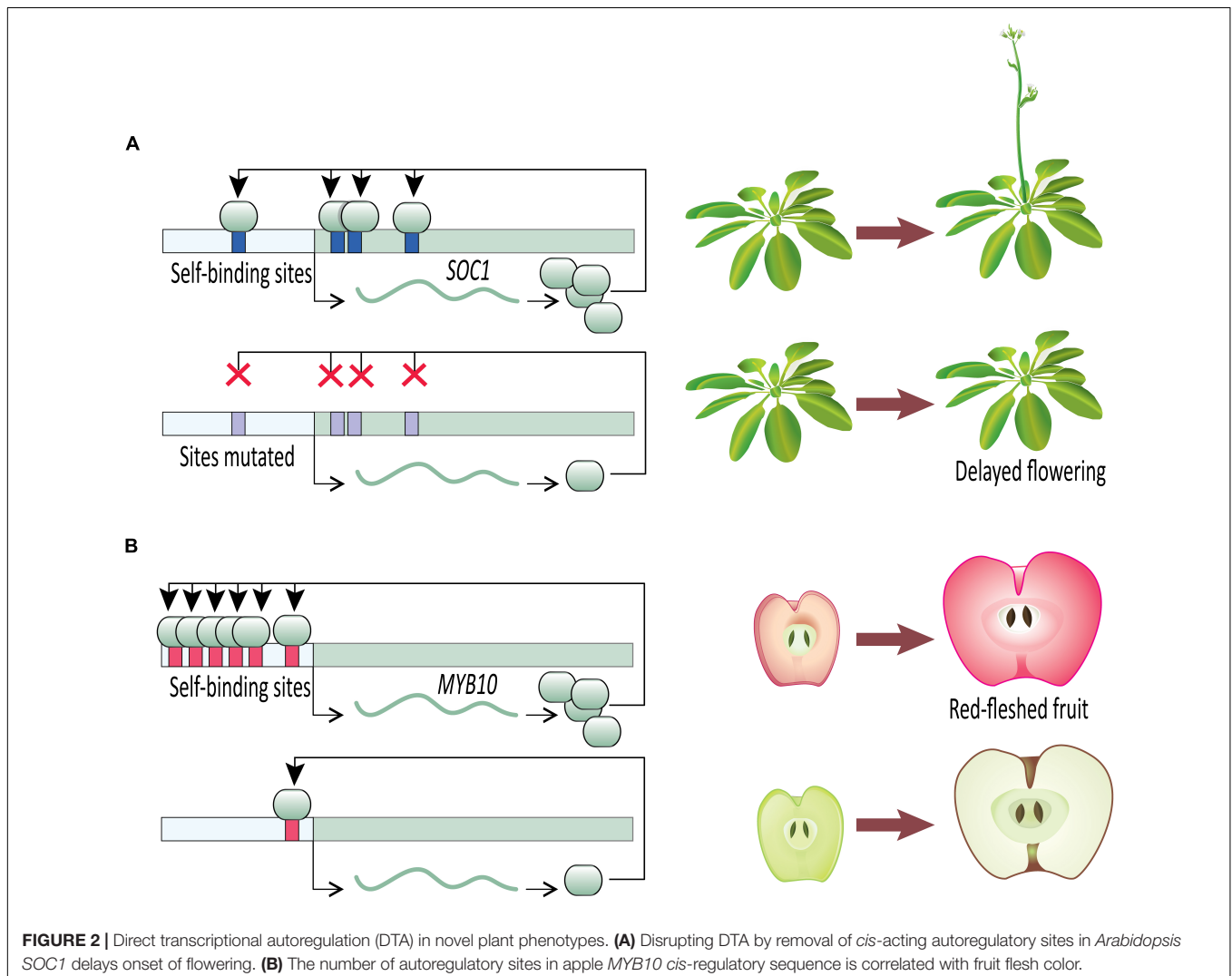


FIGURE 2 | Direct transcriptional autoregulation (DTA) in novel plant phenotypes. **(A)** Disrupting DTA by removal of *cis*-acting autoregulatory sites in *Arabidopsis* *SOC1* delays onset of flowering. **(B)** The number of autoregulatory sites in apple *MYB10* *cis*-regulatory sequence is correlated with fruit flesh color.

transcription start site (Tao et al., 2012). The flowering transition is significantly delayed in the insertional mutant *soc1-2* which carries a loss-of-function mutation in the coding sequence of *SOC1*. The delayed flowering phenotype is largely rescued when *soc1-2* lines are transformed with a wild type *SOC1* allele (including the wild type promoter). This mutant-rescue system with known self-binding sites in the *SOC1* promoter creates an elegant system for testing the specific role of *SOC1* DTA in establishing tight control of the flowering time phenotype. In heterozygous rescue lines where the self-binding sites in the transgenic allele have been mutated by substituting nucleotides at the first two and last two positions of the *CArG-box* binding site, flowering is delayed (Tao et al., 2012). This suggests that the DTA of *SOC1* has a key role in transition to flowering. Tao et al. (2012) provide further evidence of *SOC1* autoregulation using an estradiol-inducible expression system. Estradiol-induction allows tight control over transgenic protein entering the nucleus and functioning as a transcription factor. Within 2 h of estradiol-induction of transgenic *SOC1*, expression of endogenous *SOC1* tripled in comparison to a control. This rapid increase in *SOC1* expression after releasing transgenic *SOC1* protein to the nucleus suggests *SOC1* plays a direct role in its own upregulation. Together, these *SOC1* experiments in *Arabidopsis* provide clear evidence that once induced, a transcription factor undergoing DTA can rapidly increase its expression level to swiftly respond to a signal and affect developmental outcomes.

Sustained, stable, and high expression is likely key to defining complex phenotypes. Other than increasing the expression level at a certain point during development (as described in *SOC1* above), DTA would provide selective advantage if it could sustain the expression for an extended time through consecutive developmental events. A way to test this would be to determine how expression changes when homologous autoregulatory and non-autoregulatory sites between a pair of recently diverged paralogs are swapped. *Arabidopsis* *APETALA1* (*AtAP1*) and *CAULIFLOWER* (*AtCAL*) are two recently duplicated paralogs (Wang et al., 2012) and this system was employed by Ye et al. (2016) to test the role of DTA for sustaining expression in developmental patterning. *AtAP1* defines sepal development, and Ye et al. (2016) found that strong expression of *AtAP1* is initiated in floral meristems, and that the expression continues to near-mature flower stages (stage-12). *AtAP1* also undergoes DTA wherein it binds to a *CArG-box* located in its *cis*-regulatory region and activates *AtAP1* transcription. On the other hand, *AtCAL* does not undergo DTA, is expressed at a low level in early stage flowers, with the expression vanishing soon after stage-4 (Ye et al., 2016). In an elegant system, Ye et al. (2016) generated β -glucuronidase (*GUS*) reporter-constructs driven by *AtAP1* and *AtCAL* promoter regions. When the *CArG-box* in the *GUS* reporter construct with the *AtAP1* promoter was replaced with the homologous non-autoregulatory nucleotides from the *AtCAL* promoter, two changes occurred. First, the overall expression level of *GUS* dropped, and second, the expression duration was shortened, approximating that of *AtCAL* in wild type plants. On the contrary, when *GUS* was placed under the control of an *AtCAL* promoter whose non-autoregulatory nucleotides

had been replaced with the homologous *CArG-box* from the *AtAP1* promoter, *GUS* expression level increased and extended to near-mature stage flowers. This suggests that DTA of *AtAP1* not only has a role in maintaining high expression levels compared to the non-autoregulated paralog, but has a critical role in sustaining the expression for an extended period. This study did not directly test the role of *AtAP1* DTA, its loss or acquisition, in defining phenotype. However, direct evidence for acquisition or loss of DTA on the evolution of a novel phenotype comes from domesticated apples.

Malus domestica (domesticated apple) provides compelling evidence for the importance of DTA on phenotypic outcomes (Figure 2B). The color of fruit flesh in many domesticated apple varieties ranges from white to red. Variation in fruit color is regulated by the transcription factor MYB10, which upregulates anthocyanin expression, especially cyanidin-3-galactoside (Espley et al., 2007, 2009, 2013). Anthocyanin-regulating MYBs have been reported from a wide variety of angiosperm species (reviewed in Lin-Wang et al., 2010), including *Malus* (Espley et al., 2009), *Prunus* (Starkevič et al., 2015), *Myrica* (Niu et al., 2010), *Arabidopsis* (Gonzalez et al., 2008), and *Ipomoea* (Mano et al., 2007). *Malus domestica* has two alleles of MYB10 that are identical in their coding sequences but differ in their promoter sequences. Allele *R1* promoter contains one MYB10 autoregulatory binding site, whereas allele *R6* promoter contains six repeats of the autoregulatory site (Espley et al., 2009). The white-fleshed domestic apple varieties are homozygous for the one-repeat *R1* allele, whereas the red-fleshed varieties are *R1/R6* heterozygotes or *R6/R6* homozygotes, which leads to increased anthocyanin production via DTA (Espley et al., 2009). It is not clear which allele is ancestral in domesticated apples. Of the four *Malus* species that contributed to the domesticated apple genome (Cornille et al., 2014), *M. sieversii* can be either *R6/R6* (Espley et al., 2009; Lin-Wang et al., 2010) or *R1/R6* (Espley et al., 2009; van Nocker et al., 2012), and *M. baccata* is *R1/R1* (van Nocker et al., 2012). Of the other species in the genus *Malus* tested for MYB10 promoter sequence, all but one have the *R1/R1* genotype (van Nocker et al., 2012).

Though it is not clear whether *R1/R1* (white flesh) or *R6/R6* (red flesh) is ancestral in the genus *Malus*, it is clear from studies in domesticated apple that changes to fruit flesh color are regulated by addition or loss of autoregulatory sites in the MYB10 promoter. The evidence from flesh coloration in apples suggests an interesting possibility. Self-activating loops of DTA can serve as easy modules for evolving elevated or reduced gene expression levels. Such evolutionary shifts in gene expression have potentially adaptive developmental consequences accompanied by minimal pleiotropy. Genes, including transcription factors, are often regulated by *trans*-activators that bind to the *cis*-acting elements in the regulatory region of the target gene. Theoretically, these target genes can be upregulated in three ways: adding more *cis*-regulatory sites recognized by either the existing or novel *trans*-activators, upregulating the expression of the existing *trans*-activators, or acquiring new (or additional) self-binding sites in the promoter region. Addition of *cis*-regulatory sites recognized by *trans*-activators can be ineffective if the expression level of the *trans*-activator is limiting. Additionally,

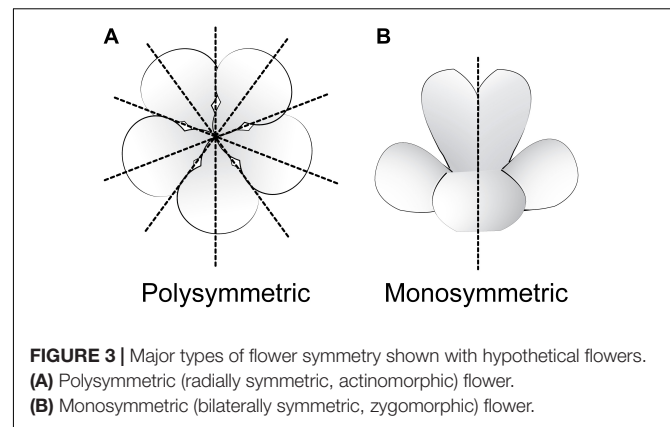
increasing the expression level of the *trans*-activator can have pleiotropic consequences. However, acquiring new (or additional) *cis*-regulatory self-binding sites can lead to increased expression of the target gene while bypassing the limitations associated with *trans*-activation. Similarly, reduced expression levels can evolve with minimal pleiotropic consequences through the loss of existing autoregulatory sites.

The evidence from SOC1, AtAP1, and MYB10 provide insight into why genes involved in defining novel phenotypes are likely to undergo DTA. Autoregulatory loops can serve as a quick developmental switch that can rapidly respond to an inbound signal, they can provide high expression levels, and extend that expression through consecutive developmental events. Lastly, DTA can act as a module that can be used to evolve increased or decreased expression with minimal pleiotropic effect, allowing the evolution of novel phenotypes that require such directional changes in protein levels. Quick evolutionary shifts in developmental function of paralogs and divergent alleles can therefore occur through gain or loss of DTA, most likely through gain or amplification of self-binding sites in *cis*-regulatory sequences of focal genes.

EVIDENCE FOR DTA IN FLOWER SYMMETRY EVOLUTION

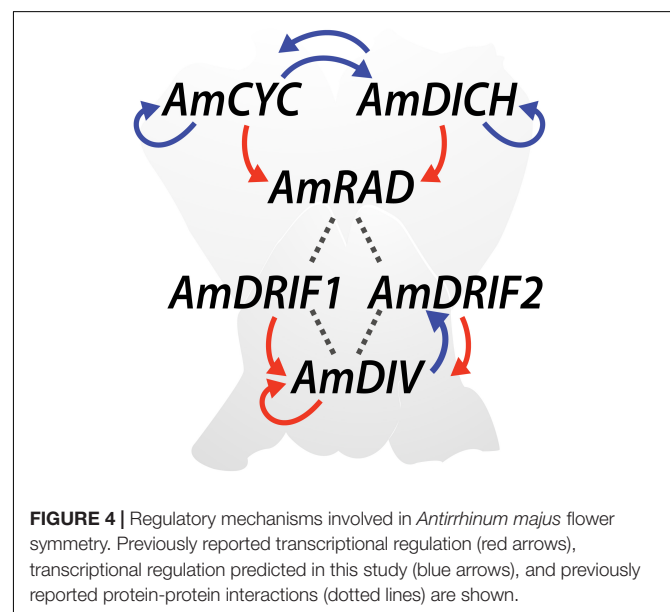
An emerging system for studying the role of DTA in both development and evolution is flower symmetry. DTA has been implicated in the control of monosymmetry (bilateral symmetry; zygomorphy) (Yang et al., 2012), and may represent a critical step for the evolution of this floral novelty. Monosymmetric flowers are considered a key innovation defining flower form in many species-rich flowering plant lineages including Lamiales, asterids, legumes, and orchids (Sargent, 2004; Vamossi and Vamossi, 2010). Therefore, assessing the role of DTA in the development of flower monosymmetry may provide critical insights into patterns of gene network modification that facilitate novel trait evolution. Below, we review the genetic control of monosymmetry in Lamiales alongside the evidence for DTA. We test for previously unreported regulatory interactions in the *Antirrhinum majus* flower symmetry program, as well as the potential for DTA in a major radiation of taxa with primarily monosymmetric flowers, the Lamiales. Lastly, we comment on possible wide-spread DTA in repeated origins of monosymmetry across flowering plants.

Flowering plants are ancestrally polysymmetric (radially symmetric; actinomorphic; **Figure 3**) (Sauquet et al., 2017). Evolutionary shifts away from polysymmetry include asymmetry (no axis of flower symmetry) and disymmetry (two non-equivalent axes of flower symmetry), but monosymmetry (a single axis of flower mirror-image symmetry; **Figure 3**) is by far the most common form of non-radial symmetry in flowering plants. Monosymmetric flowers have evolved at least 130 times independently during flowering plant diversification (Reyes et al., 2016). The role of floral symmetry in pollination was recognized as early as 1793 by Sprengel in his monumental German work *Das entdeckte Geheimniss der Natur im Bau und in der Befruchtung der Blumen* (reviewed in Neal et al., 1998;



Endress, 1999; Fenster et al., 2004, 2009). Monosymmetric flowers are often associated with specialized pollination by animals (Kampny, 1995; reviewed in Neal et al., 1998), rarely in wind pollinated species (Yuan et al., 2009), and transitions to monosymmetry are strongly associated with increased speciation rates (Sargent, 2004; O'Meara et al., 2016).

The genetics of monosymmetry is best understood in the model species *A. majus* (snapdragon, Lamiales). The flowers of *A. majus* have two distinct morphological regions—the dorsal (top; adaxial) side, and the ventral (bottom; abaxial) side (**Figure 4**). Monosymmetry of *A. majus* flowers along the dorso-ventral axis is defined by a competitive interaction involving TCP and MYB transcription factors. TCP (TEOSINTE BRANCHED1, CYCLOIDEA, and PROLIFERATING CELL FACTORS) and MYB (first described from avian myeloblastosis virus) proteins are found as large gene families in flowering plants (Yanhui et al., 2006; Martín-Trillo and Cubas, 2010) and play diverse roles in aspects of vegetative and reproductive developmental patterning (Martín-Trillo and Cubas, 2010; Ambawat et al., 2013;



Parapunova et al., 2014). The dorsal side of *A. majus* flowers is defined by the combined action of two recently duplicated TCP paralogs, *CYCLOIDEA* (*AmCYC*) and *DICHOTOMA* (*AmDICH*) (Luo et al., 1996, 1999; Hileman and Baum, 2003; Corley et al., 2005). These two transcription factors define dorsal flower morphology partly by activating the transcription of a downstream MYB protein, *RADIALIS* (*AmRAD*; **Figure 4**) (Corley et al., 2005). *AmRAD* post-translationally negatively regulates another MYB protein, *DIVARICATA* (*AmDIV*), which defines ventral flower morphology. Through this negative interaction, *AmRAD* excludes the ventral flower identity specified by *AmDIV* from the dorsal side of the developing *A. majus* flower (**Figure 4**). Specifically, *AmRAD* and *AmDIV* compete for interaction with two MYB-family protein partners called *DIV* and *RAD* Interacting Factors 1 and 2 (*AmDRIF1* and *AmDRIF2*) (Almeida et al., 1997; Galego and Almeida, 2002; Corley et al., 2005; Raimundo et al., 2013). *AmDIV* requires protein-protein interaction with *AmDRIF1* or 2 to function as a transcription factor and upregulate its own transcription, as well as to regulate downstream targets (**Figure 4**) (Perez-Rodriguez et al., 2005; Raimundo et al., 2013). In the dorsal flower domain, *AmRAD* outcompetes *AmDIV* for interaction with *AmDRIF1/2*, preventing accumulation of *AmDIV* protein (Raimundo et al., 2013).

Because flower monosymmetry has evolved multiple times, a considerable amount of effort has gone into testing whether elements of the *A. majus* symmetry program function to specify dorso-ventral differentiation in other flowering plant lineages. Interestingly, all monosymmetric species tested at a molecular level so far show evidence that a TCP-based regulatory network is likely involved in differentiation along the dorso-ventral flower axis. These studies span eudicot and monocot lineages and primarily, but not exclusively, show a pattern of dorsal-specific floral expression of *TCP* homologs (for example, Citerne et al., 2003, 2010; Busch and Zachgo, 2007; Wang et al., 2008; Yuan et al., 2009; Bartlett and Specht, 2011; Howarth et al., 2011; Chapman et al., 2012; Preston and Hileman, 2012; Damerval et al., 2013; and reviewed in Hileman, 2014). In core eudicots, there are three lineages of *CYCLOIDEA* (*CYC*)-like TCP genes resulting from two rounds of duplication near the origin of core eudicots: the *CYC1*-, *CYC2*-, and *CYC3*-lineages (Howarth and Donoghue, 2006; Citerne et al., 2013). *AmCYC* and *AmDICH* belong to the *CYC2*-lineage, and in an interesting pattern, all TCP genes implicated in floral monosymmetry in core eudicots belong to the same *CYC2*-lineage (Howarth and Donoghue, 2006; Citerne et al., 2010; and reviewed in Hileman, 2014). How these orthologous genes were recruited convergently during the multiple evolutionary origins of floral monosymmetry, from an as yet unclear function in species with ancestral polysymmetry, remains an open question.

Detailed developmental studies in *A. majus* have provided key insights into the regulatory interactions that shape flower monosymmetry, and *A. majus* as a model represents a species-rich lineage of flowering plants, Lamiales. Monosymmetry evolved early in Lamiales diversification (Zhong and Kellogg, 2015; Reyes et al., 2016), and developmental genetic studies in additional Lamiales species provide further

insight into the regulatory network that shapes bilateral flower symmetry across the entire lineage. Notably, detailed expression and functional studies of *CYC*, *RAD* and *DIV* orthologs in Gesneriaceae, a sister lineage to the bulk of Lamiales species diversity, have contributed to a fuller understanding of regulatory interactions that shape Lamiales flower monosymmetry (Citerne et al., 2000; Smith et al., 2004; Gao et al., 2008; Zhou et al., 2008; Yang et al., 2010, 2012; Liu et al., 2014a). From studies in *A. majus* (Plantaginaceae) and *Primulina heterotricha* (syn. *Chirita heterotricha*; Gesneriaceae), there is strong evidence that at least two components of the flower symmetry network undergo DTA–*DIV* and *CYC* (**Figure 4**).

As mentioned above, *AmDIV* forms heterodimers with *AmDRIF1* and 2 to specify ventral flower identity in *A. majus* (Raimundo et al., 2013). *AmDIV*–*AmDRIF* dimers bind to a consensus sequence that includes the conserved *I*-box motif, 5'-GATAAG-3' located 2596 bp upstream of the *AmDIV* transcription start site (Raimundo et al., 2013), providing compelling evidence that *AmDIV* is involved in an autoregulatory loop. Autoregulation of *DIV* orthologs has not been tested outside of *A. majus*. In *P. heterotricha*, peloric (radialized) forms due to flower ventralization have reduced expression levels of *CYC* orthologs, *PhCYC1C* and *PhCYC1D* (Yang et al., 2012), presenting strong evidence that these two genes define dorsal identity of monosymmetric *P. heterotricha* flowers. Experimental evidence suggests that *PhCYC1* and *PhCYC2* undergo DTA; *PhCYC1* and *PhCYC2* proteins bind to the consensus TCP-binding sequence 5'-GGNCCC-3' in the putative promoter regions of both *PhCYC1* and *PhCYC2* (Yang et al., 2010, 2012). Autoregulation of *CYC* orthologs has not been tested outside of *P. heterotricha*.

These initial insights from *A. majus* and *P. heterotricha* lead to a set of important evolutionary questions. Is autoregulation of *CYC* orthologs conserved across Lamiales? And has a pattern of autoregulation repeatedly evolved in *CYC2*-lineage orthologs from lineages with independently derived monosymmetric flowers? This second question is especially compelling given that *CYC2*-lineage ortholog expression is expected to persist from early through later stages of flower development in order to specify asymmetric morphological differentiation along the dorso-ventral floral axis in lineages with flower monosymmetry.

Methods: Evidence for DTA in Flower Symmetry Evolution

Homolog Predictions and Phylogenetic Analyses

AmCYC, *AmDICH*, *AmRAD*, and *AmDIV* orthologs were identified from published sources and online databases by tBLASTx (Altschul et al., 1990). The gene names/identifiers and sources are listed in **Supplementary Tables 1, 2**. Gene identifiers are also included with terminal genes on the phylogenies (**Supplementary Figures 1, 2**). A subset of included genes were available as full-length coding sequences from public databases. A subset of included genes were available as partial coding sequences from public databases. For partial coding sequences from species with available genome data, we predicted the full-length coding sequences either manually by aligning to

previously reported homologs, or by prediction with AUGUSTUS (Stanke et al., 2004). A subset of included genes were identified by BLAST (Altschul et al., 1990) from annotated genomes. We predicted the coding sequences either manually or with AUGUSTUS when our BLAST searches hit a region in a genome where no or partial genes were predicted. For *Mimulus lewisii* *DIV* and *RAD* homologs, we first BLAST searched the available transcriptome and subsequently mapped the hits to the genome. Two sets of sequences used here were not publicly available, the genes from *Ipomoea lacunosa* whose genome sequence was generously shared by Dr. Mark Rausher (Duke University), and *Mimulus guttatus* *RADlike1*, which was shared by Dr. Jinshun Zhong (University of Vermont; the sequence was reported in Zhong et al., 2017).

We translationally aligned the coding sequences (omitting the stop codon) of *CYC*-like genes using MAFFT v7.388 (Kato et al., 2002) in Geneious 10.2.3 (Kearse et al., 2012) with the following parameters: algorithm—auto, scoring matrix—BLOSUM62, gap opening penalty—1.1, offset value—0.124. The entire alignment was used for downstream phylogenetic analyses. The *CYC*-like gene tree was estimated using a Bayesian approach (Metropolis-coupled Markov chain Monte Carlo) in MrBayes 3.2.6 (Ronquist et al., 2012) with uninformative priors for 10 million generations on the online CIPRES portal at <https://www.phylo.org> (Miller et al., 2010). The core-eudicot *CYC*-like tree was rooted with Ranunculales *CYC*-like genes in FigTree¹.

DIV- and *RAD*-like genes were translationally aligned using an approach similar to *CYC*-like genes except for the following: gap opening penalty—1.53, and offset value—0.123. We removed the columns with 70% or more gaps from the alignment, and from the subsequent file used only the conserved first *MYB* domain and nucleotides immediately 3' to this domain. *DIV*- and *RAD*-like gene trees were estimated using the same approach as for *CYC*-like genes. Resulting *DIV*- and *RAD*-like trees were mid-point rooted in FigTree¹. For all sequences included in our phylogenetic analyses, nexus format nucleotide alignment along with the Bayesian parameter block, and the unaligned coding sequences in fasta format available from the Dryad Digital Repository: <https://doi.org/10.5061/dryad.tv54037>.

Consensus TCP and *DIV*-Binding Site Predictions

We downloaded up to 3 kb non-coding sequence upstream of the transcription start sites of target Lamiales *CYC*, *RAD*, and *DIV* homologs from corresponding genomes. We downloaded up to 3 kb non-coding sequence upstream of the transcription start sites of representative core eudicot *CYC* homologs from corresponding genomes. All genomic sources are listed in **Supplementary Table 1**. Within these sequences, we searched for the consensus TCP-binding site 5'-GGNCCC-3' (Kosugi and Ohashi, 2002; Costa et al., 2005; Yang et al., 2012; Gao et al., 2015) on both strands using Geneious 10.2.3 (Kearse et al., 2012). In *A. majus* only, we searched for the consensus *DIV*-binding site, 5'-[AGC]GATA[AC][GC][GAC]-3' (Raimundo et al., 2013) in 3 kb upstream non-coding sequences of the six genes known to be involved in *A. majus* flower symmetry (**Figure 4**) using

Geneious 10.2.3 (Kearse et al., 2012). To determine whether the consensus TCP-binding sites found in the *A. majus* and *M. lewisii* upstream *CYC* homolog sequences were derived from other genomic locations, we used the predicted TCP-binding sites, plus 100 bp on either side, as BLAST queries against the available genomes in Geneious 10.2.3 (Kearse et al., 2012).

Analysis of Motif Enrichment

We tested for consensus TCP-binding site enrichment using *Analysis of Motif Enrichment* (AME²; McLeay and Bailey, 2010). AME can identify known or user-provided motifs that are relatively enriched in a given set of sequences compared with shuffled versions of those sequences or with user-provided control sequences. AME does not discriminate among motifs based on their locations within the sequences. The following options were selected: sequence scoring method—average odds score, motif enrichment test—rank sum test, and background model—uniform model. We defined the consensus TCP-binding site as 5'-GGNCCC-3' (Kosugi and Ohashi, 2002; Costa et al., 2005; Yang et al., 2012; Gao et al., 2015), and query sequences as 3 kb upstream of transcription start sites of focal genes, and used shuffled sequences as the control. The upstream non-coding sequences are available in fasta format from the Dryad Digital Repository: <https://doi.org/10.5061/dryad.tv54037>.

Quantitative Reverse-Transcriptase PCR (rt-PCR)

Antirrhinum majus wild type (genotype JI 7) and *divaricata* mutants (genotype JI 13) were acquired from John Innes Centre, United Kingdom, under USDA Permit No. P37-16-01034. Five flower buds of the same developmental stage (stage-11, flower bud *ca.* 4.0 mm in length, corolla equal in length to calyx, petal tips white in wild type; Vincent and Coen, 2004) were sampled from each genotype. RNA was extracted using RNeasy plant minikit (Qiagen, Germantown, MD, United States), followed by DNase treatment (TURBOTM DNase, Thermo Fisher Scientific, Waltham, MA, United States), and cDNA synthesis (iScript cDNA Synthesis Kit, Bio-Rad, Hercules, CA, United States). Quantitative rt-PCR was performed on a StepOnePlusTM Real-Time PCR System (Thermo Fisher Scientific) using SYBRTM Select Master Mix (Thermo Fisher Scientific). Quantitative rt-PCR was carried out for three technical replicates for each of five biological replicates per genotype. Expression was normalized against *UBIQUITIN5*. This gene has been reported to have little transcriptional variation across tissue types and developmental stages (Preston and Hileman, 2010). Expression was analyzed by the $\Delta\Delta C_t$ method. Significant differences in relative expression between genotypes were determined using two sample *t*-test assuming equal variances in Minitab. The quantitative rt-PCR primers were as follows: AmDRIF1_RT_F4: GCCTTGGATCAAATTTCGGC; AmDRIF1_RT_R4: AGGAAGAATGGAGCTGGCAA; AmDRIF2_RT_F1a: AATGGTCATGGAGAGTGGGG; AmDRIF2_RT_R1: TATAGCTTGCTCCTCTGGGG; AmUBQ5_qPCR_F1: GCGC AAGAAGAAGACCTACAC; AmUBQ5_qPCR_R1: CTTCC

¹<http://tree.bio.ed.ac.uk/software/figtree/>

²<http://meme-suite.org/tools/ame>

TGAGCCTCTGCACTT. Efficiency of PCR was determined using DART (Peirson et al., 2003).

Results: Evidence for DTA in Flower Symmetry Evolution

Predicted TCP- and DIV-Binding Sites in *A. majus* Are Consistent With Known and Hypothesized Transcriptional Regulation

In *A. majus*, we found consensus TCP-binding sites in four of the six genes known to be involved in *A. majus* flower symmetry (Figure 4 and Table 1). *AmCYC* and *AmDICH* had eight and four predicted TCP-binding sites in their upstream non-coding sequences, respectively, and likely regulate their own and each other's expression. Notably, *AmCYC* DTA has been hypothesized previously (Costa et al., 2005), and the presence of predicted autoregulatory sites in *AmCYC* and *AmDICH* is consistent with the putative auto and cross-regulation of *P. heterotricha* *PhCYC1C* and *PhCYC1D* (Yang et al., 2012). *AmRAD*, known to be positively regulated by *AmCYC* and *AmDICH* (Corley et al., 2005; Costa et al., 2005), had two predicted consensus TCP-binding sites in its upstream non-coding sequence. *AmDIV* and *AmDRIF2* did not have predicted TCP-binding sites in their upstream non-coding sequences, consistent with evidence that they are unlikely to be under direct transcriptional regulation by *AmCYC*, *AmDICH*, or any other more distantly related TCP transcription factors.

Consensus TCP-binding sites (plus 100 bp flanking sequence from either side) initially identified in the upstream non-coding sequences of *AmCYC* and *AmDICH* were used to search for similar sites elsewhere in the *A. majus* genome. These searches

resulted in only self-hits to *AmCYC* and *AmDICH* upstream non-coding sequences or cross-paralog matches between *AmCYC* and *AmDICH*. This result suggests that these sites evolved *de novo* and not through translocation of existing sites from elsewhere in the genome. Similarly, our search for consensus TCP-binding sites from *M. lewisii* *CYC2*-lineage genes in the *M. lewisii* genome resulted in only self-hits.

We identified two consensus DIV-binding sites in the *AmDIV* upstream non-coding sequence (Table 2), one of which was previously implicated by Raimundo et al. (2013) in *AmDIV* DTA. *AmCYC*, *AmRAD*, *AmDRIF1* and *AmDRIF2*, but not *AmDICH*, also had predicted DIV-binding sites in their upstream non-coding sequences (Table 2). It is unlikely that the predicted DIV-binding sites in the upstream non-coding sequences of *AmCYC* or *AmRAD* function for *AmDIV* binding. This is because *AmDIV* function is impaired in the presence of *AmRAD* proteins through competitive inhibition.

Expression Analyses Suggest Additional Autoregulation of DIV in *A. majus*

Given the presence of predicted DIV-binding sites in *AmDRIF1* and *AmDRIF2* upstream non-coding sequences (Table 2), we tested whether *AmDRIF1* and/or *AmDRIF2* expression is significantly altered in the *A. majus* *div* mutant background compared to wild type. We found that *AmDRIF1*, despite having multiple *DIV* consensus binding sites in its upstream region, was not under either direct or indirect regulation by *AmDIV* ($p = 0.453$; Figure 5A). *AmDRIF1* may be regulated by a non-DIV MYB transcription factor(s) that binds to the consensus DIV-binding motif. On the other hand, we found significantly lower levels of *AmDRIF2* expression in *div* mutant flower buds compared to wild type ($p = 0.031$; Figure 5B). This suggests that *AmDRIF2* is either directly or indirectly positively regulated by *AmDIV*. In turn, *AmDIV* is positively regulated by *AmDRIF2*-*AmDIV* heterodimers (Raimundo et al., 2013).

TABLE 1 | Predicted consensus TCP-binding sites in the upstream non-coding sequences of *A. majus* flower symmetry genes.

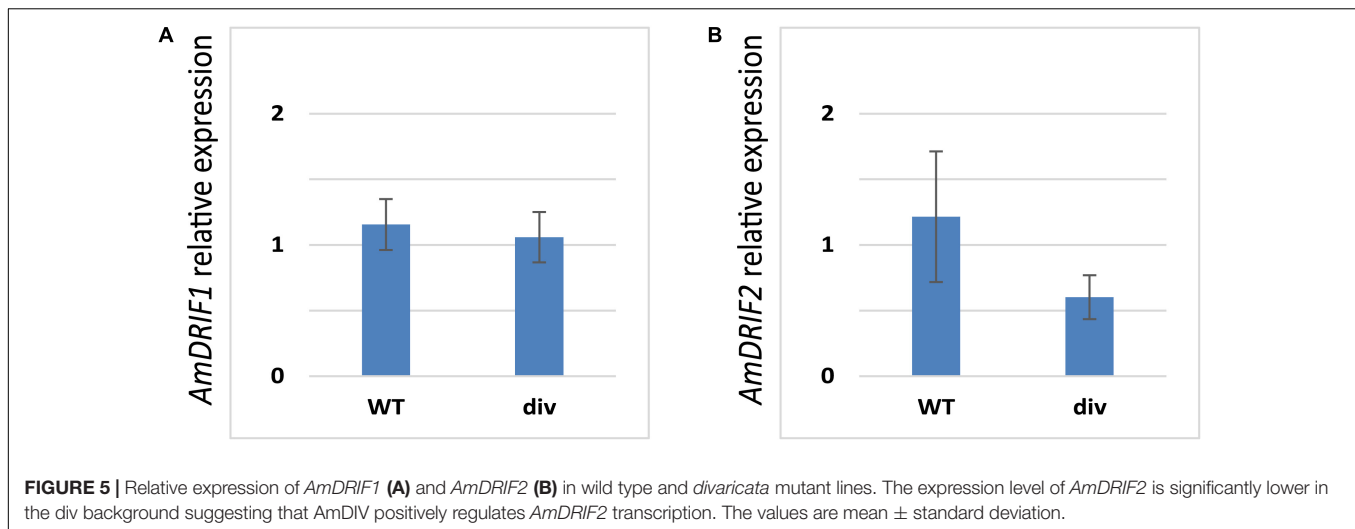
Gene	Sequence	DNA strand	bp upstream of transcription start
<i>AmCYC</i>	GGGCCC	Sense	2454–2457
	GGGCCC	Sense	544–549
	GGGCCC	Anti-sense	544–549
	GGGCCC	Anti-sense	2452–2457
	GGCCCC	Sense	2451–2456
	GGCCCC	Sense	2292–2297
	GGCCCC	Sense	543–548
	GGCCCC	Anti-sense	2453–2458
<i>AmDICH</i>	GGGCCC	Sense	1170–1175
	GGGCCC	Anti-sense	1170–1175
	GGCCCC	Sense	965–970
	GGCCCC	Anti-sense	1171–1176
<i>AmRAD</i> Costa et al., 2005	GGCCCC	Sense	1521–1526
	GGCCCC	Sense	1489–1494
<i>AmDRIF1</i>	GGTCCC	Anti-sense	2394–2399

Bases in bold indicate conservation in the consensus binding site. *AmDIV* and *AmDRIF2* lack consensus TCP-binding sites in their upstream non-coding sequences. Costa et al. (2005) reported TCP-binding sites for *AmRAD* and suggested the presence of autoregulatory sites in the non-coding sequence upstream of *AmCYC*.

TABLE 2 | Predicted consensus DIV-binding sites in the upstream non-coding sequences of *A. majus* flower symmetry genes.

Gene	Sequence	DNA strand	bp upstream of transcription start
<i>AmCYC</i>	AGATAAGG	Anti-sense	329–336
<i>AmRAD</i>	AGATAACA	Anti-sense	798–805
	GGATAACG	Anti-sense	1051–1058
	CGATAAGA	Anti-sense	2843–2850
<i>AmDIV</i> Raimundo et al., 2013	AGATAAGG	Sense	2595–2602
<i>AmDRIF1</i>	CGATACCC	Sense	1557–1564
	GGATACGG	Sense	711–718
	AGATAAGG	Sense	242–249
	AGATAAGC	Anti-sense	505–512
<i>AmDRIF2</i>	AGATAACC	Anti-sense	1892–1899

Bases in bold indicate conservation in the consensus binding site. *AmDICH* lacks consensus DIV-binding sites in upstream non-coding sequences. One of two consensus DIV-binding sites in *AmDIV* was reported by Raimundo et al. (2013).



Therefore, *AmDIV* appears to experience both direct and ITA through interaction of *AmDIV* *cis*-regulatory sequences with *AmDRIF2*-*AmDIV* heterodimers.

Putative TCP-Binding Sites Are Enriched in Upstream Non-coding Sequences of Lamiales *CYC2*-Lineage Genes

While no *CYC2*-lineage gene outside *P. heterotricha* has been experimentally tested for DTA, it is possible to infer the potential for DTA by screening for the consensus TCP-binding site, 5'-GGNCCC-3' (Kosugi and Ohashi, 2002; Yang et al., 2012; Gao et al., 2015), in putative *cis*-regulatory regions of Lamiales *CYC2*-lineage genes. Given that flower monosymmetry is homologous in *P. heterotricha* and *A. majus*, evolving early in the diversification of Lamiales (Zhong and Kellogg, 2015; Reyes et al., 2016), a straight-forward hypothesis is that *CYC2*-lineage DTA evolved early in Lamiales and has been retained in Lamiales lineages with monosymmetric flowers. Under this hypothesis, Lamiales with flower monosymmetry will retain consensus TCP-binding site(s) in putative *CYC2*-lineage *cis*-regulatory sequences. The availability of multiple Lamiales genomes (Supplementary Table 1) allowed us to begin testing the hypothesis that autoregulation is potentially conserved across Lamiales *CYC* orthologs.

We identified orthologs of *AmCYC/AmDICH* (*CYC2*-lineage genes) from genome-sequenced Lamiales plus representative core eudicots (Supplementary Figure 1 and Table 1). We identified orthologs of *AmRAD* and *AmDIV* from genome-sequenced Lamiales plus representative orthologs from sister lineages to Lamiales, Gentianales, and Solanales (Supplementary Figure 2 and Table 2). As with *P. heterotricha* and *A. majus*, recent duplication events lead to paralog complexity for *CYC2*-lineage genes (Supplementary Figure 1). We found that at least one *CYC2*-lineage gene from each core eudicot species had consensus TCP-binding sites(s) in the upstream non-coding sequence (Supplementary Tables 3, 4), with two exceptions. The only *CYC2*-lineage genes in *Vitis vinifera* (Vitales), *CYCLOIDEA-like 2a*, and *Gossypium raimondii* (Malvales), *TCP1*, had no

consensus TCP-binding sites in their upstream non-coding sequences.

We found consensus TCP-binding sites in the upstream non-coding sequences of *CYC2*-lineage genes in a wide variety of core eudicots with flowers with mono-, poly-, and dissymmetry (Supplementary Tables 3, 4). However, *prima facie*, the *CYC2*-lineage orthologs from Lamiales appeared to be enriched for consensus TCP-binding sites. We tested for enrichment of consensus TCP-binding sites in the non-coding sequences upstream of Lamiales *CYC2*-lineage genes. Additionally, we tested the upstream non-coding sequences of non-Lamiales core-eudicot *CYC2*-lineage genes, and Lamiales *RAD* and *DIV* orthologs for enrichment in consensus TCP-binding sites. We predict that *RAD* orthologs may show enrichment of the consensus TCP-binding site due to conserved regulation of *RAD* by *CYC*-like transcription factors across Lamiales, but that Lamiales *DIV* orthologs are not likely to be enriched for the consensus TCP-binding site given that there is no previous data indicating regulation of *DIV* orthologs by *CYC*-like transcription factors or other TCP proteins.

As expected, we found that the upstream non-coding sequences of Lamiales *DIV* orthologs were not significantly enriched for the consensus TCP-binding sites ($p = 0.517$; Table 3), and that the upstream non-coding sequences of Lamiales *RAD* orthologs were significantly enriched for the consensus TCP-binding site ($p = 0.0406$; Table 3). This result is consistent with *CYC*-like transcription factors acting as regulators of *RAD*, but not *DIV* across Lamiales. Strikingly, we found that the upstream non-coding sequences of *CYC2*-lineage genes in Lamiales were significantly enriched in consensus TCP-binding sites ($p = 0.0169$; Table 3) in-line with the hypothesis that *CYC* autoregulation evolved early in Lamiales, coincident with the evolution of monosymmetric flower, and has been maintained during Lamiales diversification. Notably, this pattern of enrichment appears specific to Lamiales. We tested for similar enrichment of the consensus TCP-binding site in non-Lamiales core eudicot *CYC2*-lineage orthologs and found

TABLE 3 | Results from analysis of motif enrichment (AME) tests for consensus TCP-binding sites in the upstream non-coding sequences of symmetry gene orthologs.

Test sequences (putative <i>cis</i> -regulatory regions)	Control sequences	<i>p</i> -Value	Genes surveyed	Species surveyed
Lamiales <i>DIV</i> orthologs	Shuffled test sequences	0.517	15	9
Lamiales <i>RAD</i> orthologs	Shuffled test sequences	0.0406	33	9
Lamiales <i>CYC2</i> orthologs	Shuffled test sequences	0.0169	20	9
Non-Lamiales core eudicot <i>CYC2</i> orthologs	Shuffled test sequences	0.352	39	17

Significant *p*-values (below 0.05) are in bold.

no evidence for a similar pattern of binding site enrichment ($p = 0.352$; Table 3).

DISCUSSION

Binding Site Enrichment Supports the Hypothesis That DTA of *CYC* Is Associated With the Origin of Flower Monosymmetry in Lamiales

Positive regulation of *RAD* by *CYC2*-lineage genes for specifying flower monosymmetry is conserved across much of Lamiales (Corley et al., 2005; Zhou et al., 2008; Su et al., 2017). That we find significant enrichment of consensus TCP-binding sites in Lamiales *RAD* upstream non-coding sequences is in-line with conservation of this *CYC*-*RAD* regulatory module. Strikingly, our data demonstrate that Lamiales *CYC2*-lineage genes are also significantly enriched for consensus TCP-binding sites in upstream non-coding sequences. This supports the hypothesis that the origin of Lamiales flower monosymmetry coincides with the evolution of *CYC2*-lineage DTA. Further empirical studies in emerging Lamiales models (e.g., Liu et al., 2014b; Su et al., 2017) will allow this hypothesis to be tested, as well as the alternative, that *CYC2*-lineage genes undergo transcriptional regulation by other TCP family proteins. As additional eudicot genomes become available, tests for TCP-binding site enrichment can be carried out in other lineages with bilaterally symmetrical flowers for which a role of *CYC2*-lineage genes has been implicated, for example, Fabaceae (Wang et al., 2008; Xu et al., 2013) and Malpighiaceae (Zhang et al., 2010).

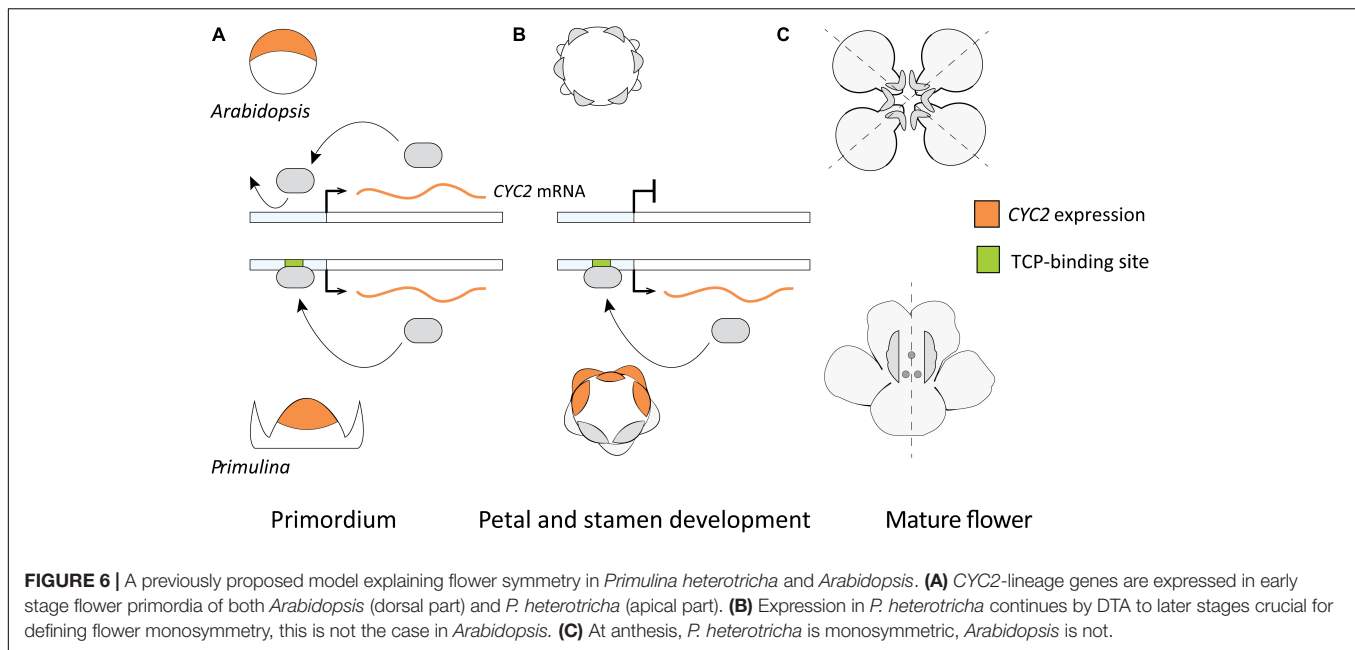
Evaluating the Pan-Eudicot Model for Monosymmetry Involving DTA of *CYC2*-Lineage Genes

A model hypothesizing the role of DTA for the parallel origin of monosymmetric flowers across eudicots was put forward by Yang et al. (2012; Figure 6) based on two primary lines of evidence. First, the observed differences in duration of flower specific expression of *CYC2*-lineage genes between species with monosymmetric vs. non-monosymmetric flowers. Second, the reported absence of consensus TCP-binding sites in the upstream non-coding sequences of *CYC2*-lineage genes from non-monosymmetric flowers. Specifically, *Arabidopsis thaliana*, *Brassica rapa*, *Vitis vinifera*, and *Solanum lycopersicum* do not have monosymmetric flowers and were reported to lack

consensus TCP-binding sites in their *CYC2*-lineage genes compared to *Glycine max*, *Medicago trunculata*, *Mimulus guttatus*, *Primulina heterotricha*, *Oryza sativa*, and *Zea mays* (representing three independent origins of monosymmetry) that have consensus TCP-binding sites (Yang et al., 2012).

This model relies heavily on observations from *Arabidopsis* flowers where the expression of the sole *CYC2*-lineage gene (*AtTCP1*) is transiently dorsal-specific and the flowers are non-monosymmetric (Cubas et al., 2001). It is clear that *AtTCP1* does not play a critical role in floral organ differentiation in *Arabidopsis*, given no floral-specific DTA or other means by which expression can persist to later stages of flower differentiation. However, the pattern in *Arabidopsis* may not be universal for non-monosymmetric flowers. Closely related monosymmetric and non-monosymmetric Brassicaceae flowers do not exhibit a consistent pattern of early dorsal-specific expression (Busch et al., 2012). Evidence from Brassicaceae suggests that *Arabidopsis*-like dorsal-restricted expression early in flower development is not a pre-requisite for the evolution of flower monosymmetry via DTA. Beyond Brassicaceae, there are examples of ancestrally non-monosymmetric flowers in core-eudicots where expression of *CYC2*-lineage genes is not localized spatially and/or restricted to an early developmental stage. These examples include *Bergia texana* (Elatinaceae) (Zhang et al., 2010), *Viburnum plicatum* (Adoxaceae) (Howarth et al., 2011), and *Solanum lycopersicum* (Solanaceae, ancestral state ambiguous) (Parapunova et al., 2014), as well as an early-diverging eudicot, *Eschscholzia californica* (Papaveraceae) (Kölsch and Gleissberg, 2006).

Yang et al. (2012) reported a correlation between flower monosymmetry vs. non-monosymmetry and the presence vs. absence of consensus TCP-binding sites in corresponding upstream non-coding sequences of *CYC2*-lineage genes. This contributed to the model for the origin of flower monosymmetry facilitated by the evolution of *CYC2*-lineage DTA. In our expanded sampling we find that consensus TCP-binding sites are present in the upstream non-coding sequences of many *CYC2*-lineage genes across eudicots irrespective of flower symmetry. Yet, in an interesting pattern, all species with independently derived monosymmetric flowers that we investigated (Fabales, Lamiales, Brassicales, Asterales) have at least one *CYC2*-lineage ortholog with a consensus TCP-binding sequence in the upstream non-coding sequences (Supplementary Tables 3, 4). On the other hand, many species with non-monosymmetric flowers also have at least one *CYC2*-lineage ortholog with a consensus TCP-binding sequence



in their upstream non-coding sequences (**Supplementary Tables 3, 4**). Notably, we find that the sole *CYC2*-lineage gene in *Arabidopsis* (*AtTCP1*), and a second *CYC2*-lineage gene in tomato that was not included in Yang et al. (2012), *Solanum lycopersicum* *TCP26* (Solyc03g045030.1), have consensus TCP-binding sites in their upstream non-coding sequences (**Supplementary Table 4**).

AtTCP1 binds to all combinations of the consensus sequence 5'-GGNCCC-3' *in vitro*, and flanking regions have limited significance in this interaction (Gao et al., 2015). *In vivo*, *AtTCP1* can directly bind to the two TCP-binding sites located in the regulatory region of the downstream gene *DWARF4* (Gao et al., 2015). This suggests that the *Arabidopsis* *TCP1* transcription factor can likely bind to the predicted TCP-binding site in its own upstream non-coding sequence, and hence possibly undergoes DTA. *AtTCP1* is expressed and is functional across the shoot organs throughout development, from seedlings to inflorescences (Koyama et al., 2010). This persistent expression is consistent with it having a predicted autoregulatory site. Expression surveys employing *in situ* mRNA hybridization (Cubas et al., 2001) and *AtTCP1* promoter fused to a β -glucuronidase (GUS) construct (Koyama et al., 2010) did not detect *AtTCP1* expression in later stages of flower development. It is interesting that the expression of a gene that is widely expressed in and controls development of many different organs is specifically downregulated in flowers. It is possible that *AtTCP1* is negatively regulated during late stages of *Arabidopsis* flower development, or continues to be expressed in flowers but a level that can only be detected by more sensitive methods, like quantitative rt-PCR.

Predicted *CYC2*-lineage autoregulatory sites are strongly associated with monosymmetry supporting the potential importance for DTA in establishing high and continuous asymmetric expression through later stages of flower organ differentiation (**Figure 6**). However, this pattern is not

exclusive: *CYC2*-lineage orthologs from many species lacking monosymmetry also have predicted TCP-binding sites. This may be autoregulation for alternative developmental pathways, or regulation of *CYC2*-lineage genes by upstream TCP activators. At this point, experimental tests of TCP gene autoregulation are too sparse to draw solid conclusions regarding the role of DTA in independent origins of flower monosymmetry across core eudicots.

Origin and Evolution of Autoregulatory Sites in DTA

Any *cis*-regulatory site can evolve by two primary processes, *de novo* by mutation and/or recombination in ancestral non-regulatory sequences, or by duplication of existing regulatory sites from a different location in the genome. Both have been reported in the origin of *cis*-regulatory sites involved in DTA. For example, the *CAR*-box sites involved in *Arabidopsis* *AP1* autoregulation discussed earlier evolved by substitutions in the ancestral sequence that likely had a weak affinity for *AP1* (Ye et al., 2016). Once evolved, these sites can undergo duplications, as reported in the apple *MYB10* gene that controls fruit flesh color (Espley et al., 2009; van Nocker et al., 2012).

How did the predicted autoregulatory sites in *CYC2*-lineage genes originate? We did not detect consensus TCP-binding sites with accompanying flanking sequences elsewhere in the *A. majus* or *M. lewisii* genomes. This suggests that these predicted autoregulatory sites evolved *in situ* and are not a result of duplication from a different part of the genome, i.e., similar to the origin of the autoregulatory sites in *Arabidopsis* *AP1* (Ye et al., 2016). However, multiple consensus TCP-binding sites are present within single *A. majus* and *M. lewisii* *CYC2*-lineage genes. To further test whether these multiple TCP-binding sites within a single putative regulatory region evolved by local,

intra-genic duplication, as in the case of *MYB10* promoter in apples (Espley et al., 2009; van Nocker et al., 2012), we aligned all *A. majus* and *M. lewisii* consensus TCP-binding sites, along with 100 bp flanking on either side, from within single upstream non-coding regions. We found no evidence that any of the predicted TCP-binding sites are derived from tandem duplication within *CYC* regulatory regions, again suggesting that multiple binding sites evolved *de novo*.

Chicken or Egg: Novel Function or DTA First?

We have discussed potential roles of DTA in development, but how does DTA itself evolve? Autoregulation is common among genes positioned upstream in genetic regulatory networks with crucial developmental functions (discussed in Crews and Pearson, 2009; Hoot et al., 2010; specifically tested in yeasts and hepatocytes by Pasqualucci et al., 2003; Odom et al., 2006; Hervay et al., 2011; Tao et al., 2012). This observed pattern leads to an interesting chicken or egg conundrum. Which evolves first in genes recruited to new developmental functions: the novel function, or the autoregulation? Two scenarios can explain the observed pattern that crucial genes are often autoregulated. (1) DTA evolves first, and such genes are recruited for new functions that require extended stable expression. Or, (2) New function evolves first, and such genes, under selective pressure to provide extended stable expression, evolve DTA.

Evidence supporting scenario 2 is found in the *Arabidopsis AtAP1* example. This A-class floral homeotic gene in Brassicaceae underwent a duplication that generated the paralogs *API* and *CAL* gene lineages (Wang et al., 2012). *AtAP1* defines sepals in *Arabidopsis thaliana*, but this function has not been reported elsewhere, and is likely an innovation in the genus *Arabidopsis* (Huijser et al., 1992; Lowman and Purugganan, 1999; Shepard and Purugganan, 2002; Litt, 2007; Ruokolainen et al., 2010). Except for the *API* paralog in *Arabidopsis* species, no Brassicaceae *API/CAL* gene tested to date undergoes DTA (Ye et al., 2016). And, as described above, DTA is an integral component of *AtAP1* A-class function in flower development. Further, while the *API* orthologs of two *Arabidopsis* species have *CAR-G-box* in their *cis*-regulatory region that allows them to undergo DTA, other Brassicaceae species have *CAR-G-box*-like sequences with mismatches in the homologous gene region. In one such homolog, *Capsella rubella API*, the binding affinity of the mismatched *CAR-G-box*-like sequence was tested and can only weakly bind to *API* protein. Hence, *Capsella rubella API* is likely not autoregulated (Ye et al., 2016). This suggests that the autoregulation of *Arabidopsis API* evolved either after or during, but not before, its recruitment to A-class function.

A major unanswered question that will clarify the origin of DTA in *Arabidopsis API* is whether its orthologs have similar functions in other Brassicaceae species. It is challenging to identify the ancestral state of autoregulation for any gene primarily for two reasons: there has been little functional work outside the model species, and predictive surveys are limited

because genomes sequencing has been biased toward lineages with those model species. As plant sciences expands away from model systems (Poaceae, Brassicaceae, and Solanaceae), a wider phylogenetic sampling will facilitate reconstruction of ancestral molecular interactions.

CONCLUSION

The origins and evolution of autoregulation will likely remain elusive until extensive experimental evidence emerges from multiple plant (and animal) lineages that inform ancestral and derived roles for autoregulation in development. It is, however, not surprising that a large number of transcription factors involved in defining crucial or novel phenotypes undergo DTA, as this form of regulation is expected to both enhance and stabilize gene expression patterns critical for developmental patterning. We find evidence for enrichment of self-binding sites in Lamiales *CYC2*-lineages genes. This enrichment may reflect evolution of a novel pattern of DTA early in Lamiales diversification, coincident with the origin of a key morphological innovation, floral monosymmetry. It is likely that the putative autoregulatory binding sites associated with Lamiales *CYC2*-lineages genes evolved via *de novo* mutations. Whether DTA is conserved across Lamiales awaits further experimental evidence, as does the hypothesis that independent origins of flower monosymmetry may be associated with the evolution of positive transcriptional autoregulation.

AUTHOR CONTRIBUTIONS

LH co-conceived of this project, oversaw analyses, and contributed to writing the manuscript. AS co-conceived of this project, carried out analyses, and contributed to writing the manuscript.

FUNDING

This work was supported by Ecology and Evolutionary Biology General Research Fund and The Botany Endowment at The University of Kansas.

ACKNOWLEDGMENTS

The authors thank members of the Hilleman lab for insightful discussions and Dr. Mark Rausher for sharing an early draft of the *Ipomoea lacunosa* genome.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.01561/full#supplementary-material>

REFERENCES

- Almeida, J., Rocheta, M., and Galego, L. (1997). Genetic control of flower shape in *Antirrhinum majus*. *Development* 124, 1387–1392.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Ambawat, S., Sharma, P., Yadav, N. R., and Yadav, R. C. (2013). MYB transcription factor genes as regulators for plant responses: an overview. *Physiol. Mol. Biol. Plants* 19, 307–321. doi: 10.1007/s12298-013-0179-1
- Bartlett, M. E., and Specht, C. D. (2011). Changes in expression pattern of the *TEOSINTE BRANCHED1*-like genes in the Zingiberales provide a mechanism for evolutionary shifts in symmetry across the order. *Am. J. Bot.* 98, 227–243. doi: 10.3732/ajb.1000246
- Bateman, E. (1998). Autoregulation of eukaryotic transcription factors. *Prog. Nucleic Acid Res. Mol. Biol.* 60, 133–168. doi: 10.1016/S0079-6603(08)60892-2
- Busch, A., Horn, S., Mühlhausen, A., Mummenhoff, K., and Zachgo, S. (2012). Corolla monosymmetry: evolution of a morphological novelty in the brassicaceae family. *Mol. Biol. Evol.* 29, 1241–1254. doi: 10.1093/molbev/msr297
- Busch, A., and Zachgo, S. (2007). Control of corolla monosymmetry in the Brassicaceae *Iberis amara*. *Proc. Natl. Acad. Sci. U.S.A.* 104, 16714–16719. doi: 10.1073/pnas.0705338104
- Chapman, M. A., Tang, S., Draeger, D., Nambeesan, S., Shaffer, H., Barb, J. G., et al. (2012). Genetic analysis of floral symmetry in Van Gogh's sunflowers reveals independent recruitment of *CYCLOIDEA* genes in the Asteraceae. *Public Lib. Sci. Genet.* 8:e1002628. doi: 10.1371/journal.pgen.1002628
- Citerne, H., Jabbour, F., Nadot, S., and Damerval, C. (2010). The evolution of floral symmetry. *Adv. Bot. Res.* 54, 85–137.
- Citerne, H. L., Le Guilloux, M., Sannier, J., Nadot, S., and Damerval, C. (2013). Combining phylogenetic and syntenic analyses for understanding the evolution of TCP ECE genes in Eudicots. *PLoS One* 8:e74803. doi: 10.1371/journal.pone.0074803
- Citerne, H. L., Luo, D., Pennington, R. T., Coen, E., and Cronk, Q. C. B. (2003). A phylogenomic investigation of *CYCLOIDEA*-Like TCP genes in the Leguminosae. *Plant Physiol.* 131, 1042–1053. doi: 10.1104/pp.102.016311
- Citerne, H. L., Möller, M., and Cronk, Q. C. B. (2000). Diversity of *cycloidea*-like Genes in gesneriaceae in relation to floral symmetry. *Ann. Bot.* 86, 167–176. doi: 10.1006/anbo.2000.1178
- Corley, S. B., Carpenter, R., Copsey, L., and Coen, E. (2005). Floral asymmetry involves an interplay between TCP and MYB transcription factors in *Antirrhinum*. *Proc. Natl. Acad. Sci. U.S.A.* 102, 5068–5073. doi: 10.1073/pnas.0501340102
- Cornille, A., Giraud, T., Smulders, M. J. M., Roldán-Ruiz, I., and Gladieux, P. (2014). The domestication and evolutionary ecology of apples. *Trends Genet.* 30, 57–65. doi: 10.1016/j.tig.2013.10.002
- Costa, M. M. R., Fox, S., Hanna, A. I., Baxter, C., and Coen, E. (2005). Evolution of regulatory interactions controlling floral asymmetry. *Development* 132, 5093–5101. doi: 10.1242/dev.02085
- Crews, S. T., and Pearson, J. C. (2009). Transcriptional autoregulation in development. *Curr. Biol.* 19, R241–R246. doi: 10.1016/j.cub.2009.01.015
- Cripps, R. M., Lovato, T. L., and Olson, E. N. (2004). Positive autoregulation of the *Myocyte enhancer factor-2* myogenic control gene during somatic muscle development in *Drosophila*. *Dev. Biol.* 267, 536–547. doi: 10.1016/j.ydbio.2003.12.004
- Cubas, P., Coen, E., and Zapater, J. M. M. (2001). Ancient asymmetries in the evolution of flowers. *Curr. Biol.* 11, 1050–1052. doi: 10.1016/S0960-9822(01)00295-0
- Damerval, C., Citerne, H., Guilloux, M. L., Domenichini, S., Duthiel, J., Ronse de Craene, L., et al. (2013). Asymmetric morphogenetic cues along the transverse plane: shift from disymmetry to zygomorphy in the flower of Fumarioideae. *Am. J. Bot.* 100, 391–402. doi: 10.3732/ajb.1200376
- Endress, P. K. (1999). Symmetry in flowers: diversity and evolution. *Int. J. Plant Sci.* 160, S3–S23.
- Espley, R. V., Bovy, A., Bava, C., Jaeger, S. R., Tomes, S., Norling, C., et al. (2013). Analysis of genetically modified red-fleshed apples reveals effects on growth and consumer attributes. *Plant Biotechnol. J.* 11, 408–419. doi: 10.1111/pbi.12017
- Espley, R. V., Brendolise, C., Chagné, D., Kutty-Amma, S., Green, S., Volz, R., et al. (2009). Multiple repeats of a promoter segment causes transcription factor autoregulation in red apples. *Plant Cell* 21, 168–183. doi: 10.1105/tpc.108.059329
- Espley, R. V., Hellens, R. P., Putterill, J., Stevenson, D. E., Kutty-Amma, S., and Allan, A. C. (2007). Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10. *Plant J.* 49, 414–427. doi: 10.1111/j.1365-3113X.2006.02964.x
- Fenster, C. B., Armbruster, W. S., and Dudash, M. R. (2009). Specialization of flowers: is floral orientation an overlooked first step? *New Phytol.* 183, 502–506. doi: 10.1111/j.1469-8137.2009.02852.x
- Fenster, C. B., Armbruster, W. S., Wilson, P., Dudash, M. R., and Thomson, J. D. (2004). Pollination syndromes and floral specialization. *Annu. Rev. Ecol. Evol. Syst.* 35, 375–403. doi: 10.1146/annurev.ecolsys.34.011802.132347
- Galego, L., and Almeida, J. (2002). Role of DIVARICATA in the control of dorsoventral asymmetry in *Antirrhinum* flowers. *Genes Dev.* 16, 880–891. doi: 10.1101/gad.221002
- Gallo-Ebert, C., Donigan, M., Liu, H.-Y., Pascual, F., Mannes, M., Pandya, D., et al. (2013). The yeast anaerobic response element AR1b regulates aerobic antifungal drug-dependent sterol gene expression. *J. Biol. Chem.* 288, 35466–35477. doi: 10.1074/jbc.M113.526087
- Gao, Q., Tao, J.-H., Yan, D., Wang, Y.-Z., and Li, Z.-Y. (2008). Expression differentiation of CYC-like floral symmetry genes correlated with their protein sequence divergence in *Chirita heterotricha* (Gesneriaceae). *Dev. Genes Evol.* 218, 341–351. doi: 10.1007/s00427-008-0227-y
- Gao, Y., Zhang, D., and Li, J. (2015). TCP1 modulates DWF4 expression via directly interacting with the GGNGCC Motifs in the promoter region of DWF4 in *Arabidopsis thaliana*. *J. Genet. Genomics* 42, 383–392. doi: 10.1016/j.jgg.2015.04.009
- Gonzalez, A., Zhao, M., Leavitt, J. M., and Lloyd, A. M. (2008). Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in *Arabidopsis* seedlings. *Plant J. Cell Mol. Biol.* 53, 814–827. doi: 10.1111/j.1365-3113X.2007.03373.x
- Hervay, N. T., Hodurova, Z., Balazfyova, Z., and Gbelska, Y. (2011). Autoactivated KIPDR1 gene in the control of multidrug resistance in *Kluyveromyces lactis*. *Can. J. Microbiol.* 57, 844–849. doi: 10.1139/w11-071
- Hileman, L. C. (2014). Bilateral flower symmetry — how, when and why? *Curr. Opin. Plant Biol.* 17, 146–152. doi: 10.1016/j.pbi.2013.12.002
- Hileman, L. C., and Baum, D. A. (2003). Why do paralogs persist? Molecular evolution of *CYCLOIDEA* and related floral symmetry genes in Antirrhineae (Veroniceae). *Mol. Biol. Evol.* 20, 591–600. doi: 10.1093/molbev/msg063
- Hiratsu, K., Matsui, K., Koyama, T., and Ohme-Takagi, M. (2003). Dominant repression of target genes by chimeric repressors that include the EAR motif, a repression domain, in *Arabidopsis*. *Plant J.* 34, 733–739. doi: 10.1046/j.1365-3113X.2003.01759.x
- Hochschild, A. (2002). The λ switch: cI closes the gap in autoregulation. *Curr. Biol.* 12, R87–R89. doi: 10.1016/S0960-9822(02)00667-X
- Holloway, D. M., Lopes, F. J., da Fountoura Costa, F., Travençolo, B. A., Golyandina, N., Usevich, K., et al. (2011). Gene expression noise in spatial patterning: hunchback promoter structure affects noise amplitude and distribution in *Drosophila* segmentation. *PLoS Comput. Biol.* 7:e1001069. doi: 10.1371/journal.pcbi.1001069
- Hoot, S. J., Brown, R. P., Oliver, B. G., and White, T. C. (2010). The *UPC2* promoter in *Candida albicans* contains two cis-acting elements that bind directly to Upc2p, resulting in transcriptional autoregulation. *Eukaryot. Cell* 9, 1354–1362. doi: 10.1128/EC.00130-10
- Howarth, D. G., and Donoghue, M. J. (2006). Phylogenetic analysis of the “ECE” (CYC/TB1) clade reveals duplications predating the core eudicots. *Proc. Natl. Acad. Sci. U.S.A.* 103, 9101–9106. doi: 10.1073/pnas.0602827103
- Howarth, D. G., Martins, T., Chimney, E., and Donoghue, M. J. (2011). Diversification of *CYCLOIDEA* expression in the evolution of bilateral flower symmetry in Caprifoliaceae and Lonicera (Dipsacales). *Ann. Bot.* 107, 1521–1532. doi: 10.1093/aob/mcr049
- Huijser, P., Klein, J., Lönning, W. E., Meijer, H., Saedler, H., and Sommer, H. (1992). Bracteomania, an inflorescence anomaly, is caused by the loss of function of the MADS-box gene *squamosa* in *Antirrhinum majus*. *EMBO J.* 11, 1239–1249.
- Kampny, C. M. (1995). Pollination and flower diversity in Scrophulariaceae. *Bot. Rev.* 61, 350–366.

- 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
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199
- Kölsch, A., and Gleissberg, S. (2006). Diversification of *CYCLOIDEA*-like TCP Genes in the basal Eudicot families Fumariaceae and Papaveraceae s.str. *Plant Biol.* 8, 680–687. doi: 10.1055/s-2006-924286
- Kosugi, S., and Ohashi, Y. (2002). DNA binding and dimerization specificity and potential targets for the TCP protein family. *Plant J.* 30, 337–348. doi: 10.1046/j.1365-3113.2002.01294.x
- Koyama, T., Sato, F., and Ohme-Takagi, M. (2010). A role of TCP1 in the longitudinal elongation of leaves in *Arabidopsis*. *Biosci. Biotechnol. Biochem.* 74, 2145–2147. doi: 10.1271/bbb.100442
- Lin-Wang, K., Bolitho, K., Grafton, K., Kortstee, A., Karunairatnam, S., McGhie, T. K., et al. (2010). An R2R3 MYB transcription factor associated with regulation of the anthocyanin biosynthetic pathway in Rosaceae. *BMC Plant Biol.* 10:50. doi: 10.1186/1471-2229-10-50
- Litt, A. (2007). An evaluation of A-function: evidence from the *APETALA1* and *APETALA2* gene lineages. *Int. J. Plant Sci.* 168, 73–91. doi: 10.1086/509662
- Liu, B.-L., Pang, H.-B., Yang, X., and Wang, Y.-Z. (2014a). Functional and evolutionary analyses of *Primulina heterotricha* *CYC1C* gene in tobacco and *Arabidopsis* transformation systems. *J. Syst. Evol.* 52, 112–123. doi: 10.1111/jse.12067
- Liu, B.-L., Yang, X., Liu, J., Dong, Y., and Wang, Y.-Z. (2014b). Characterization, efficient transformation and regeneration of *Chirita pumila* (Gesneriaceae), a potential evo-devo model plant. *Plant Cell Tissue Organ Cult.* 118, 357–371. doi: 10.1007/s11240-014-0488-2
- Lowman, A. C., and Purugganan, M. D. (1999). Duplication of the *Brassica oleracea* *APETALA1* floral homeotic gene and the evolution of domesticated cauliflower. *J. Hered.* 90, 514–520.
- Luo, D., Carpenter, R., Copsey, L., Vincent, C., Clark, J., and Coen, E. (1999). Control of organ asymmetry in flowers of *Antirrhinum*. *Cell* 99, 367–376. doi: 10.1016/S0092-8674(00)81523-8
- Luo, D., Carpenter, R., Vincent, C., Copsey, L., and Coen, E. (1996). Origin of floral asymmetry in *Antirrhinum*. *Nature* 383, 794–799. doi: 10.1038/383794a0
- Ma, X., Zhu, Q., Chen, Y., and Liu, Y.-G. (2016). CRISPR/Cas9 platforms for genome editing in plants: developments and applications. *Mol. Plant* 9, 961–974. doi: 10.1016/j.molp.2016.04.009
- Mano, H., Ogasawara, F., Sato, K., Higo, H., and Minobe, Y. (2007). Isolation of a regulatory gene of anthocyanin biosynthesis in tuberous roots of purple-fleshed sweet potato. *Plant Physiol.* 143, 1252–1268. doi: 10.1104/pp.106.094425
- Martínez-Antonio, A., and Collado-Vides, J. (2003). Identifying global regulators in transcriptional regulatory networks in bacteria. *Curr. Opin. Microbiol.* 6, 482–489. doi: 10.1016/j.mib.2003.09.002
- Martin-Trillo, M., and Cubas, P. (2010). TCP genes: a family snapshot ten years later. *Trends Plant Sci.* 15, 31–39. doi: 10.1016/j.tplants.2009.11.003
- McLeay, R. C., and Bailey, T. L. (2010). Motif enrichment analysis: a unified framework and an evaluation on ChIP data. *BMC Bioinformatics* 11:165. doi: 10.1186/1471-2105-11-165
- Miller, M. A., Pfeiffer, W., and Schwartz, T. (2010). “Creating the CIPRES Science Gateway for inference of large phylogenetic trees,” in *Proceedings of the 2010 Gateway Computing Environments Workshop (GCE)* (New Orleans, LA: IEEE), 1–8. doi: 10.1109/GCE.2010.5676129
- Neal, P. R., Dafni, A., and Giurfa, M. (1998). Floral symmetry and its role in plant-pollinator systems: terminology, distribution, and hypotheses. *Annu. Rev. Ecol. Syst.* 29, 345–373. doi: 10.1146/annurev.ecolsys.29.1.345
- Niu, S.-S., Xu, C.-J., Zhang, W.-S., Zhang, B., Li, X., Lin-Wang, K., et al. (2010). Coordinated regulation of anthocyanin biosynthesis in Chinese bayberry (*Myrica rubra*) fruit by a R2R3 MYB transcription factor. *Planta* 231, 887–899. doi: 10.1007/s00425-009-1095-z
- Odom, D. T., Dowell, R. D., Jacobsen, E. S., Nekudova, L., Rolfe, P. A., Danford, T. W., et al. (2006). Core transcriptional regulatory circuitry in human hepatocytes. *Mol. Syst. Biol.* 2:2006.0017. doi: 10.1038/msb410.0059
- O’Maoiléidigh, D. S., Graciet, E., and Wellmer, F. (2014). Gene networks controlling *Arabidopsis thaliana* flower development. *New Phytol.* 201, 16–30. doi: 10.1111/nph.12444
- O’Meara, B. C., Smith, S. D., Armbruster, W. S., Harder, L. D., Hardy, C. R., Hileman, L. C., et al. (2016). Non-equilibrium dynamics and floral trait interactions shape extant angiosperm diversity. *Proc. R. Soc. B Biol. Sci.* 283:20152304. doi: 10.1098/rspb.2015.2304
- Parapunova, V., Busscher, M., Busscher-Lange, J., Lammers, M., Karlova, R., Bovy, A. G., et al. (2014). Identification, cloning and characterization of the tomato TCP transcription factor family. *BMC Plant Biol.* 14:157. doi: 10.1186/1471-2229-14-157
- Pasqualucci, L., Migliazza, A., Basso, K., Houldsworth, J., Chaganti, R. S. K., and Dalla-Favera, R. (2003). Mutations of the *BCL6* proto-oncogene disrupt its negative autoregulation in diffuse large B-cell lymphoma. *Blood* 101, 2914–2923. doi: 10.1182/blood-2002-11-3387
- Peirson, S. N., Butler, J. N., and Foster, R. G. (2003). Experimental validation of novel and conventional approaches to quantitative real-time PCR data analysis. *Nucleic Acids Res.* 31:e73. doi: 10.1093/nar/gng073
- Perez-Rodriguez, M., Jaffe, F. W., Butelli, E., Glover, B. J., and Martin, C. (2005). Development of three different cell types is associated with the activity of a specific MYB transcription factor in the ventral petal of *Antirrhinum majus* flowers. *Development* 132, 359–370. doi: 10.1242/dev.01584
- Pinho, R., Garcia, V., Irimia, M., and Feldman, M. W. (2014). Stability depends on positive autoregulation in boolean gene regulatory networks. *PLoS Comput. Biol.* 10:e1003916. doi: 10.1371/journal.pcbi.1003916
- Preston, J. C., and Hileman, L. C. (2010). SQUAMOSA-PROMOTER BINDING PROTEIN 1 initiates flowering in *Antirrhinum majus* through the activation of meristem identity genes. *Plant J. Cell Mol. Biol.* 62, 704–712. doi: 10.1111/j.1365-3113.2010.04184.x
- Preston, J. C., and Hileman, L. C. (2012). Parallel evolution of TCP and B-class genes in Commelinaceae flower bilateral symmetry. *EvoDevo* 3:6. doi: 10.1186/2041-9139-3-6
- Raimundo, J., Sobral, R., Bailey, P., Azevedo, H., Galego, L., Almeida, J., et al. (2013). A subcellular tug of war involving three MYB-like proteins underlies a molecular antagonism in *Antirrhinum* flower asymmetry. *Plant J.* 75, 527–538. doi: 10.1111/tpj.12225
- Reyes, E., Sauquet, H., and Nadot, S. (2016). Perianth symmetry changed at least 199 times in angiosperm evolution. *Taxon* 65, 945–964. doi: 10.12705/655.1
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029
- Rosenfeld, N., Elowitz, M. B., and Alon, U. (2002). Negative autoregulation speeds the response times of transcription networks. *J. Mol. Biol.* 323, 785–793. doi: 10.1016/S0022-2836(02)00994-4
- Ruokolainen, S., Ng, Y. P., Broholm, S. K., Albert, V. A., Elomaa, P., and Teeri, T. H. (2010). Characterization of SQUAMOSA-like genes in *Gerbera hybrida*, including one involved in reproductive transition. *BMC Plant Biol.* 10:128. doi: 10.1186/1471-2229-10-128
- Sargent, R. D. (2004). Floral symmetry affects speciation rates in angiosperms. *Proc. R. Soc. Lond. B Biol. Sci.* 271, 603–608. doi: 10.1098/rspb.2003.2644
- Sauquet, H., von Balthazar, M., Magallón, S., Doyle, J. A., Endress, P. K., Bailes, E. J., et al. (2017). The ancestral flower of angiosperms and its early diversification. *Nat. Commun.* 8:16047. doi: 10.1038/ncomms16047
- Shepard, K. A., and Purugganan, M. D. (2002). The genetics of plant morphological evolution. *Curr. Opin. Plant Biol.* 5, 49–55. doi: 10.1016/S1369-5266(01)00227-8
- Singh, A., and Hespanha, J. P. (2009). Evolution of gene auto-regulation in the presence of noise. *IET Syst. Biol.* 3, 368–378. doi: 10.1049/iet-syb.2009.0002
- Smith, J. F., Hileman, L. C., Powell, M. P., and Baum, D. A. (2004). Evolution of GCYC, a Gesneriaceae homolog of *CYCLOIDEA*, within Gesnerioideae (Gesneriaceae). *Mol. Phylogenet. Evol.* 31, 765–779. doi: 10.1016/j.ympev.2003.09.012
- Stanke, M., Steinkamp, R., Waack, S., and Morgenstern, B. (2004). AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res.* 32, W309–W312. doi: 10.1093/nar/gkh379

- Starkeviè, P., Paukštytė, J., Kazanavičiūtė, V., Denkovskienė, E., Stanys, V., Bendokas, V., et al. (2015). Expression and anthocyanin biosynthesis-modulating potential of sweet cherry (*Prunus avium* L.) MYB10 and bHLH Genes. *PLoS One* 10:e0126991. doi: 10.1371/journal.pone.0126991
- Su, S., Xiao, W., Guo, W., Yao, X., Xiao, J., Ye, Z., et al. (2017). The CYCLOIDEA-RADIALIS module regulates petal shape and pigmentation, leading to bilateral corolla symmetry in *Torenia fournieri* (Linderniaceae). *New Phytol.* 215, 1582–1593. doi: 10.1111/nph.14673
- Tao, Z., Shen, L., Liu, C., Liu, L., Yan, Y., and Yu, H. (2012). Genome-wide identification of SOC1 and SVP targets during the floral transition in *Arabidopsis*. *Plant J.* 70, 549–561. doi: 10.1111/j.1365-3113.2012.04919.x
- Vamosi, J. C., and Vamosi, S. M. (2010). Key innovations within a geographical context in flowering plants: towards resolving Darwin's abominable mystery. *Ecol. Lett.* 13, 1270–1279. doi: 10.1111/j.1461-0248.2010.01521.x
- van Nocker, S., Berry, G., Najdowski, J., Michelutti, R., Luffman, M., Forsline, P., et al. (2012). Genetic diversity of red-fleshed apples (*Malus*). *Euphytica* 185, 281–293. doi: 10.1007/s10681-011-0579-7
- Vincent, C. A., and Coen, E. S. (2004). A temporal and morphological framework for flower development in *Antirrhinum majus*. *Can. J. Bot.* 82, 681–690. doi: 10.1139/b04-042
- Wang, B., Zhang, N., Guo, C.-C., Xu, G.-X., Kong, H.-Z., and Shan, H.-Y. (2012). Evolutionary divergence of the APETALA1 and CAULIFLOWER proteins. *J. Syst. Evol.* 50, 502–511. doi: 10.1111/j.1759-6831.2012.00211.x
- Wang, Z., Luo, Y., Li, X., Wang, L., Xu, S., Yang, J., et al. (2008). Genetic control of floral zygomorphy in pea (*Pisum sativum* L.). *Proc. Natl. Acad. Sci.* 105, 10414–10419. doi: 10.1073/pnas.0803291105
- Xu, S., Luo, Y., Cai, Z., Cao, X., Hu, X., Yang, J., et al. (2013). Functional Diversity of CYCLOIDEA-like TCP genes in the control of zygomorphic flower development in *Lotus japonicus*. *J. Integr. Plant Biol.* 55, 221–231. doi: 10.1111/j.1744-7909.2012.01169.x
- Yang, X., Cui, H., Yuan, Z.-L., and Wang, Y.-Z. (2010). Significance of consensus CYC-binding sites found in the promoters of both *ChCYC* and *ChRAD* genes in *Chirita heterotricha* (Gesneriaceae). *J. Syst. Evol.* 48, 249–256. doi: 10.1111/j.1759-6831.2010.00086.x
- Yang, X., Pang, H.-B., Liu, B.-L., Qiu, Z.-J., Gao, Q., Wei, L., et al. (2012). Evolution of double positive autoregulatory feedback loops in CYCLOIDEA2 clade genes is associated with the origin of floral zygomorphy. *Plant Cell* 24, 1834–1847. doi: 10.1105/tpc.112.099457
- Yanhui, C., Xiaoyuan, Y., Kun, H., Meihua, L., Jigang, L., Zhaofeng, G., et al. (2006). The MYB transcription factor superfamily of Arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family. *Plant Mol. Biol.* 60, 107–124. doi: 10.1007/s11103-005-2910-y
- Ye, L., Wang, B., Zhang, W., Shan, H., and Kong, H. (2016). Gains and Losses of Cis-regulatory elements led to divergence of the Arabidopsis APETALA1 and CAULIFLOWER duplicate genes in the time, space, and level of expression and regulation of one paralog by the other. *Plant Physiol.* 171, 1055–1069. doi: 10.1104/pp.16.00320
- Yuan, Z., Gao, S., Xue, D.-W., Luo, D., Li, L.-T., Ding, S.-Y., et al. (2009). RETARDED PALEA1 controls palea development and floral zygomorphy in rice. *Plant Physiol.* 149, 235–244. doi: 10.1104/pp.108.128231
- Zhang, W., Kramer, E. M., and Davis, C. C. (2010). Floral symmetry genes and the origin and maintenance of zygomorphy in a plant-pollinator mutualism. *Proc. Natl. Acad. Sci. U.S.A.* 107, 6388–6393. doi: 10.1073/pnas.0910155107
- Zhong, J., and Kellogg, E. A. (2015). Stepwise evolution of corolla symmetry in CYCLOIDEA2-like and RADIALIS-like gene expression patterns in Lamiales. *Am. J. Bot.* 102, 1260–1267. doi: 10.3732/ajb.1500191
- Zhong, J., Preston, J. C., Hileman, L. C., and Kellogg, E. A. (2017). Repeated and diverse losses of corolla bilateral symmetry in the Lamiaceae. *Ann. Bot.* 119, 1211–1223. doi: 10.1093/aob/mcx012
- Zhou, X.-R., Wang, Y.-Z., Smith, J. F., and Chen, R. (2008). Altered expression patterns of TCP and MYB genes relating to the floral developmental transition from initial zygomorphy to actinomorphy in *Bournea* (Gesneriaceae). *New Phytol.* 178, 532–543. doi: 10.1111/j.1469-8137.2008.02384.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 © 2018 Sengupta and Hileman. 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.



Duplication and Diversification of *REPLUMLESS* – A Case Study in the Papaveraceae

Cecilia Zumajo-Cardona^{1,2}, Natalia Pabón-Mora³ and Barbara A. Ambrose^{1*}

¹ New York Botanical Garden, Bronx, NY, United States, ² The Graduate Center, City University of New York, New York, NY, United States, ³ Instituto de Biología, Universidad de Antioquia, Medellín, Colombia

OPEN ACCESS

Edited by:

Stefan de Folter,
Centro de Investigación y de Estudios
Avanzados (CINVESTAV), Mexico

Reviewed by:

David Smyth,
Monash University, Australia
Gerardo Acosta-Garcia,
Technological Institute of Celaya,
Mexico

*Correspondence:

Barbara A. Ambrose
bambrose@nybg.org

Specialty section:

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

Received: 10 August 2018

Accepted: 26 November 2018

Published: 12 December 2018

Citation:

Zumajo-Cardona C,
Pabón-Mora N and Ambrose BA
(2018) Duplication and Diversification
of *REPLUMLESS* – A Case Study
in the Papaveraceae.
Front. Plant Sci. 9:1833.
doi: 10.3389/fpls.2018.01833

There is a vast amount of fruit morphological diversity in terms of their texture, the number of carpels, if those carpels are fused or not and how fruits open to disperse the seeds. *Arabidopsis thaliana*, a model eudicot, has a dry bicarpellate silique, when the fruit matures, the two valves fall apart through the dehiscence zone leaving the seeds attached to the remaining medial tissue, called the replum. Proper replum development in *A. thaliana* is mediated by *REPLUMLESS* (*RPL*), a TALE Homeodomain protein. *RPL* represses the valve margin genetic program and the downstream dehiscence zone formation in the medial tissue of the siliques and *RPL* orthologs have conserved roles across the Brassicaceae eudicots. A *RPL* homolog, *qSH1*, has been studied in rice, a monocot, and plays a role in fruit shedding making it difficult to predict functional evolution of this gene lineage across angiosperms. Although *RPL* orthologs have been identified across all angiosperms, expression and functional analyses are scarce. In order to fill the phylogenetic gap between the Brassicaceae and monocots we have characterized the expression patterns of *RPL* homologs in two poppies with different fruit types, *Bocconia frutescens* with operculate valvate dehiscence and a persistent medial tissue, similar to a replum, and *Papaver somniferum*, a poppy with persistent medial tissue in between the multicarpellate gynoeceia. We found that *RPL* homologs in Papaveraceae have broad expression patterns during plant development; in the shoot apical meristem, during flowering transition and in many floral organs, especially the carpels. These patterns are similar to those of *RPL* in *A. thaliana*. However, our results suggest that *RPL* does not have conserved roles in the maintenance of medial persistent tissues of fruits but may be involved with establishing the putative dehiscence zone in dry poppy fruits.

Keywords: basal eudicots, *Bocconia frutescens*, fruit development, Papaveraceae, *Papaver somniferum*, *REPLUMLESS*, replum

INTRODUCTION

The *Arabidopsis thaliana* (*Arabidopsis*) gynoeceium is composed of two congenitally fused carpels, which after fertilization develop into a dry dehiscent fruit, known as a silique. This fruit is formed by two valves, which during dehiscence, separate from the replum by the tension created against the rigid lignified layer (Roeder and Yanofsky, 2006). The replum was originally described as the tissue

where the seeds remain attached after the two valves fall apart but is now described as only the outer or abaxial portion and does not include the inner septum (Ferrandiz et al., 1999; Alvarez and Smyth, 2002). The gene regulatory network involved in Arabidopsis fruit development has been extensively described (Ferrandiz et al., 1999; Ripoll et al., 2011; Reyes-Olalde et al., 2013; Chávez-Montes et al., 2015). One of the genes involved in proper replum development is *REPLUMLESS* (*RPL*; Roeder et al., 2003). *RPL* belongs to the TALE class of Homeodomain proteins with a TALE motif within the triple helix of the Homeodomain (HD) but are characterized from other TALE-HD proteins by a ZIBEL motif (Bürglin, 1997; Becker et al., 2002; Kumar et al., 2007; Mukherjee et al., 2009). Comprehensive analyses of *RPL* related sequences have found that these TALE proteins are closely related to the BELL proteins (Mukherjee et al., 2009), therefore they are also called BELL-like Homeodomain proteins (BLH; Chan et al., 1998; Becker et al., 2002; Roeder et al., 2003; Hake et al., 2004). *RPL* has broad expression patterns during *A. thaliana* development, with the highest expression levels detected in the stems, and in the replum beginning early in floral development (Byrne et al., 2003; Roeder et al., 2003; Dinnyen et al., 2005; Kanrar et al., 2006; Yu et al., 2009; Avino et al., 2012; Khan et al., 2012, 2015; Chung et al., 2013; Arnaud and Pautot, 2014; Andrés et al., 2015). The *rpl* mutant, as the name suggests is defective in replum development in the fruit, however, mutants of this gene also have vegetative defects (Roeder et al., 2003). *rpl* (also known as *pennywise*, *bellringer*, and *vaamana*) shows partial loss of apical dominance, shorter plants and defects in phyllotaxy (Byrne et al., 2003; Roeder et al., 2003; Smith and Hake, 2003; Bhatt et al., 2004). *RPL* maintains meristem identity by maintaining cell proliferation and repressing lateral organ boundary genes such as *BLADE-ON-PETIOLE1/2* (Khan et al., 2015). Moreover, during late fruit development *RPL* is restricted to the replum and negatively regulates *SHATERPROOF*, a MADS-box gene involved in the specification of the dehiscence zone (Roeder et al., 2003; Kramer et al., 2004; Fourquin and Ferrandiz, 2012). Meanwhile, *RPL* is also directly repressed by *APETALA2* (*AP2*), a protein that belongs to the *AP2/ERF* transcription factor family which is upstream of the entire fruit developmental network (Ripoll et al., 2011). *RPL* restricts valve and valve margin development and therefore is indirectly involved in proper replum formation (Alonso-Cantabrana et al., 2007).

REPLUMLESS orthologs have been identified across all angiosperms and are the result of a duplication event before angiosperm diversification that also gave rise to its sister clade *POUND FOOLISH* (*PNF*; Pabón-Mora et al., 2014). However, expression and functional studies are scarce outside Arabidopsis. In *Lepidium* species, also in the Brassicaceae, *RPL* expression is found only in leaves, at the tip of the inflorescence meristem and in developing flowers while absent from older flowers or in fruits (Mühlhausen et al., 2013). In *Oryza sativa* (rice), *RPL* appears to be one of the genes involved in its domestication. At maturity, wild rice disperses the fruit with the seed inside to guarantee propagation while, in domesticated rice the fruit remains attached to the plant to make harvest easy and increase production (Lin et al., 2007; Arnaud et al., 2011; Meyer and Purugganan, 2013). The domesticated rice phenotype is the

result of a mutation in the promoter of *Seed Shattering in Chromosome 1* (*qSH1*, the *RPL* homolog in rice) which controls the formation of the abscission layer at the base of the sterile bract (Konishi et al., 2006; Lin et al., 2007; Gasser and Simon, 2011). Available functional data suggest that *RPL* genes play different roles in Brassicaceae and monocots during flower and fruit development thus, comparative data is needed in order to assess their expression and functional evolution across angiosperms.

Here, we investigate the expression patterns of *RPL* homologs in the Papaveraceae, as they are members of the basal eudicots and exhibit different strategies for seed dispersal. Using fruit diversity within this family is key to understanding the mechanisms involved in defining the medial zone in different dry dehiscent fruit morphologies. Fruit diversity in Papaveraceae includes dry dehiscent fruits with poricidal capsules with pores coinciding with locule number as in *Papaver* (Sárkány and Szalay, 1964; Roth, 1977; Gunn, 1980; Kapoor, 1995) (Figures 1A,B), schizocarps as in *Platystemon* (Figure 1C), pores that extend basipetally leaving “baskets” full of seeds, as in *Argemone* (Figure 1D), fruits with complete longitudinal dehiscence with a remaining septum as in *Eschscholzia* (Becker et al., 2005) (Figure 1F) or in *Dicentra* (Figure 1G) and fruits with opercular dehiscence where the two valves, derived from the two carpels, fall apart from a remaining ring-like commissural tissue where the seed remains attached as in *Bocconia* (Zumajo-Cardona et al., 2017) (Figure 1E). The type of fruit dehiscence found in *Bocconia* resembles the dehiscence of the Arabidopsis fruit.

The fact that some species within Papaveraceae (i.e., *B. frutescens*) resemble the Arabidopsis silique, will allow us to understand if the same genetic network has been co-opted at different evolutionary points for similar types of fruit. In addition, the Papaveraceae belong to the order Ranunculales, which form a well-supported clade placed as the sister group to the core eudicots (Kim et al., 2004; Angiosperm Phylogeny Group, 2016) and therefore, occupy a key phylogenetic position outside of the well studied core eudicots and monocots (Drea et al., 2007). Here, we describe the expression patterns of *RPL* homologs in two poppies: *P. somniferum* and *B. frutescens* in order to assess the putative role of *RPL* genes in different fruit types in basal eudicots. This in turn will help us understand the shifts that the *RPL* gene lineage has undergone during angiosperm evolution.

MATERIALS AND METHODS

Gene Homolog Searches and Phylogenetic Analysis

We performed BLASTN targeted searches in previously assembled transcriptomes of *Bocconia frutescens* (Papaveraceae) (Arango-Ocampo et al., 2016; Zumajo-Cardona et al., 2017), using as query sequences those previously reported for other Papaveraceae *RPL* homologs [i.e., *Papaver somniferum* *RPL* (KKCW-2001151, OneKP) and *Sanguinaria canadensis* *RPL* (XHKT-2009137, OneKP; Pabón-Mora et al., 2014)] as well as the *Arabidopsis thaliana* sequence (AT5G02030).

To assess the phylogenetic position of the *Bocconia frutescens* homologs, we included *BofrRPL1*, 2 and 3 in a matrix consisting

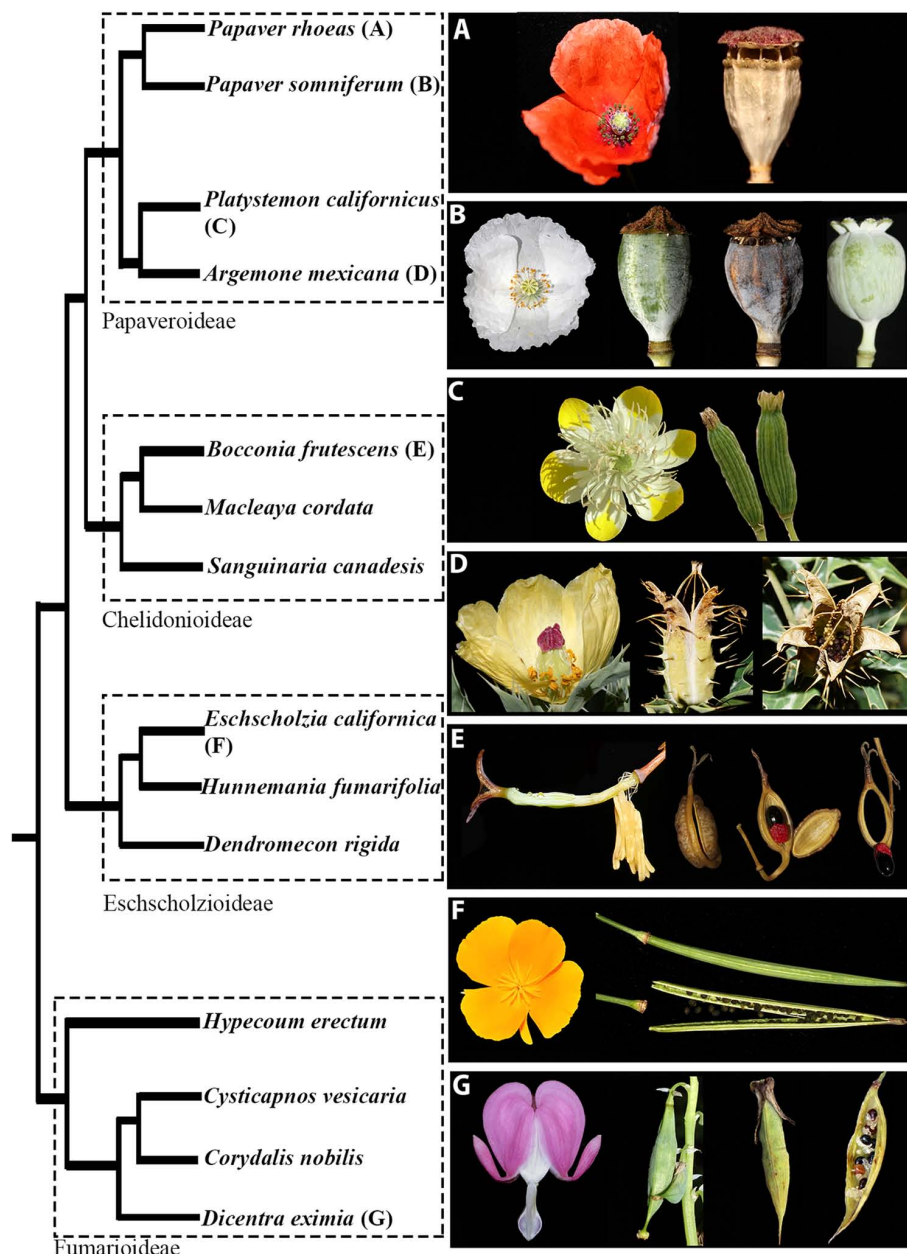


FIGURE 1 | The relationships of selected Papaveraceae with representatives from the four subfamilies: Papaveroideae, Chelidonioideae, Eschscholioideae, and Fumarioideae. Pictures of representative species illustrate the flower and fruit morphological diversity within the family. Poricidal capsules in (A) *Papaver rhoeas* and (B) *Papaver somniferum*. Note that *P. somniferum* cv. Persian White (to the right in B) lacks pores. Capsules that separate into mericarps in (C) *Platystemon californicus*. Longitudinally dehiscent capsules in, (D) *Argemone mexicana*, (F) *Eschscholzia californica* and (G) *Dicentra eximia*. Capsules with opercular dehiscence in (E) *Bocconia frutescens*.

of selected *RPL* homologs from all major plant groups, based on the sampling done by Pabón-Mora et al. (2014). Additionally, we extended the sampling particularly in basal eudicots from the plant transcriptome repositories of the oneKP database¹ and PhytoMetaSyn².

¹<https://sites.google.com/a/ualberta.ca/onekp/>

²<https://bioinformatics.tugraz.at/phytometasyn/>

A total of 117 sequences from all major angiosperm groups were compiled and edited manually to exclusively keep the open reading frame for all transcripts using AliView (Larsson, 2014). Nucleotide sequences were subsequently aligned using the online version of MAFFT³ (Katoh et al., 2002) with a gap open penalty of 3.0, offset value of 0.5, and all other

³<https://mafft.cbrc.jp/alignment/server/>

default settings. The resulting alignment was refined manually using AliView. Maximum Likelihood (ML) phylogenetic analysis using the complete nucleotide alignment of all homologs was performed through the CIPRES Science Gateway (Miller et al., 2010) with RaxML-HPC2 BlackBox (Stamatakis et al., 2008). Three *PNF* sequences from *A. thaliana*, *A. lyrata* and *C. rubella* (Brassicaceae) were used as the outgroup (**Supplementary Table S1**). Trees were observed and edited using FigTree v 1.4.3⁴. Newly isolated sequences from our *Aristolochia fimbriata* and *B. frutescens* transcriptomes are available under GenBank numbers MK057522 – MK057526. To detect conserved motifs, 28 complete protein sequences from Ranunculales were selected. The sequences were permanently translated and uploaded to the online MEME suite⁵ (Bailey et al., 2006) and run using all the default options set to find 20 motifs.

Carpel and Fruit Morphology and Anatomy

P. somniferum seeds were germinated in a growth chamber under controlled conditions with 15 h of light and a relative humidity of 60%. After germination, plants were grown to maturity under the same conditions. Flowers in preanthesis, anthesis and fruits at several developmental stages were collected and fixed in formaldehyde-acetic acid-ethanol (FAA; 3.7% formaldehyde: 5% glacial acetic acid: 50% ethanol). *B. frutescens* was collected in the field (voucher: Colombia, Antioquia, Medellín, Las Palmas, Envigado, sobre la via principal, Km 12 retorno No. 10. May 2015, C. Zumajo-Cardona and N. Pabón-Mora 03) and immediately fixed in FAA. The material was dehydrated through an alcohol-histochoice series and embedded in Paraplast X-tra (Fisher Healthcare, Houston, TX, United States). The samples were sectioned at 10 µm with a MICROM HM355 (Fisher Scientific, Pittsburgh, PA, United States) rotary microtome. Sections were stained with Johansen's safranin, to identify lignification and presence of cuticle, and 0.5% Astra Blue (Kraus et al., 1998) and mounted in Permount (Fisher Scientific, Pittsburgh, PA, United States). Sections were viewed on a Zeiss optical microscope and digitally photographed with a Nikon DXM1200C digital camera and ACT-1 software. In addition, comparative morphological analyses between the Papaveraceae fruits was done based on fresh material as shown in **Figure 1**.

In situ Hybridization Expression Analyses

P. somniferum and *B. frutescens* vegetative apices, inflorescences, floral buds and fruits at different developmental stages were collected, fixed in cold FAA and processed similarly as described above for anatomy samples. Paraplast X-tra embedded samples were maintained at 4°C until use. Samples were sectioned with a rotary microtome (Microm HM355) at 8 µm. DNA templates for RNA probe synthesis were obtained by PCR amplification of 300–370 bp fragments. To ensure specificity, the probe templates were designed to amplify the 3' sequence flanking the Homeodomain (**Supplementary Table S2** and **Supplementary Figure S1**). Fragments were cleaned

using QIAquick PCR purification Kit (Qiagen, Valencia, CA, United States). Digoxigenin labeled RNA probes were prepared using T7 RNA polymerase (Roche, Switzerland), RNase inhibitor RNasin (New England Biolabs, Ipswich, MA, United States) and RNA labeling-mix (Roche, Switzerland) according to the manufacturer's protocol. RNA *in situ* hybridization was performed according to Ambrose et al. (2000). There are a minimum of three replicates and up to nine for each probe and each developmental stage within each species. *In situ* hybridized sections were subsequently dehydrated and permanently mounted in Permount (Fisher, Waltham, MA, United States). All sections were digitally photographed using a Zeiss Axioplan microscope equipped with a Nikon DXM1200C digital camera.

RESULTS

REPLUMLESS Gene Evolution

To reconstruct the *RPL* gene lineage evolution, we used the complete coding sequence of 117 homologs from all major angiosperm groups (**Supplementary Table S1**). The sister clade *POUNDFOOLISH* (Pabón-Mora et al., 2014) was used as the outgroup. Maximum Likelihood analysis recovered independent duplication events before the diversification of Poaceae (BS = 100) and Solanaceae (BS = 100) as has been previously reported (**Figure 2**) (Pabón-Mora et al., 2014; Ortiz-Ramírez et al., 2018). Here, we recovered an additional duplication likely predating the diversification of Ranunculales (basal eudicots). We have named the two resulting clades as *RanRPL1* and *RanRPL2* (**Figure 2**). However, it is unclear if the duplication occurred before or after the radiation of Eupteleaceae, as there is a single *RPL* sequence from *Euptelea pleiosperma* in the *RanRPL1* clade. In addition, the *RanRPL1* clade includes Papaveraceae sequences and a single sequence of Menispermaceae (BS = 90) (**Figure 2**). The *RanRPL2* clade (BS = 93) includes Papaveraceae, Berberidaceae, Menispermaceae, and Ranunculaceae sequences. Multiple *RPL* sequences were identified for *Tinospora cordifolia* (Menispermaceae), *Hydrastis canadensis*, *Nigella sativa* and *Xanthorhiza simplicissima* (Ranunculaceae; **Figure 2**) but the topology does not allow us to determine if these sequences are the result of an additional duplication event predating Menispermaceae and Ranunculaceae. Finally, taxon-specific duplications in the *RPL* clade have occurred multiple times, usually associated with recent whole genome duplication (WGD) events as in the case of *Bocconia frutescens*, *Glycine max*, *Malus domestica*, *Papaver bracteatum*, *Theobroma cacao*, and *Tinospora cordifolia* (**Figure 2**) (Davie, 1935; Sugiura, 1940; Schmutz et al., 2010; Argout et al., 2011; Jain and Prasad, 2014; Panchy et al., 2016).

All of the *RPL* proteins included in the phylogenetic analysis contained the Homeodomain, the BELL domain as well as the SKY and ZIBEL motifs already characterized as highly conserved across *RPL* homologs (**Figure 3**) (Mukherjee et al., 2009; Pabón-Mora et al., 2014). The ZIBEL motif can be found before the SKY motif (i.e., BofrRPL1, 2, and 3, ChmaRPL2, PsomRPL), or after the Homeodomain (i.e., GlfrRPL1, SacaRPL,

⁴<http://tree.bio.ed.ac.uk/software/figtree/>

⁵<http://meme-suite.org/>

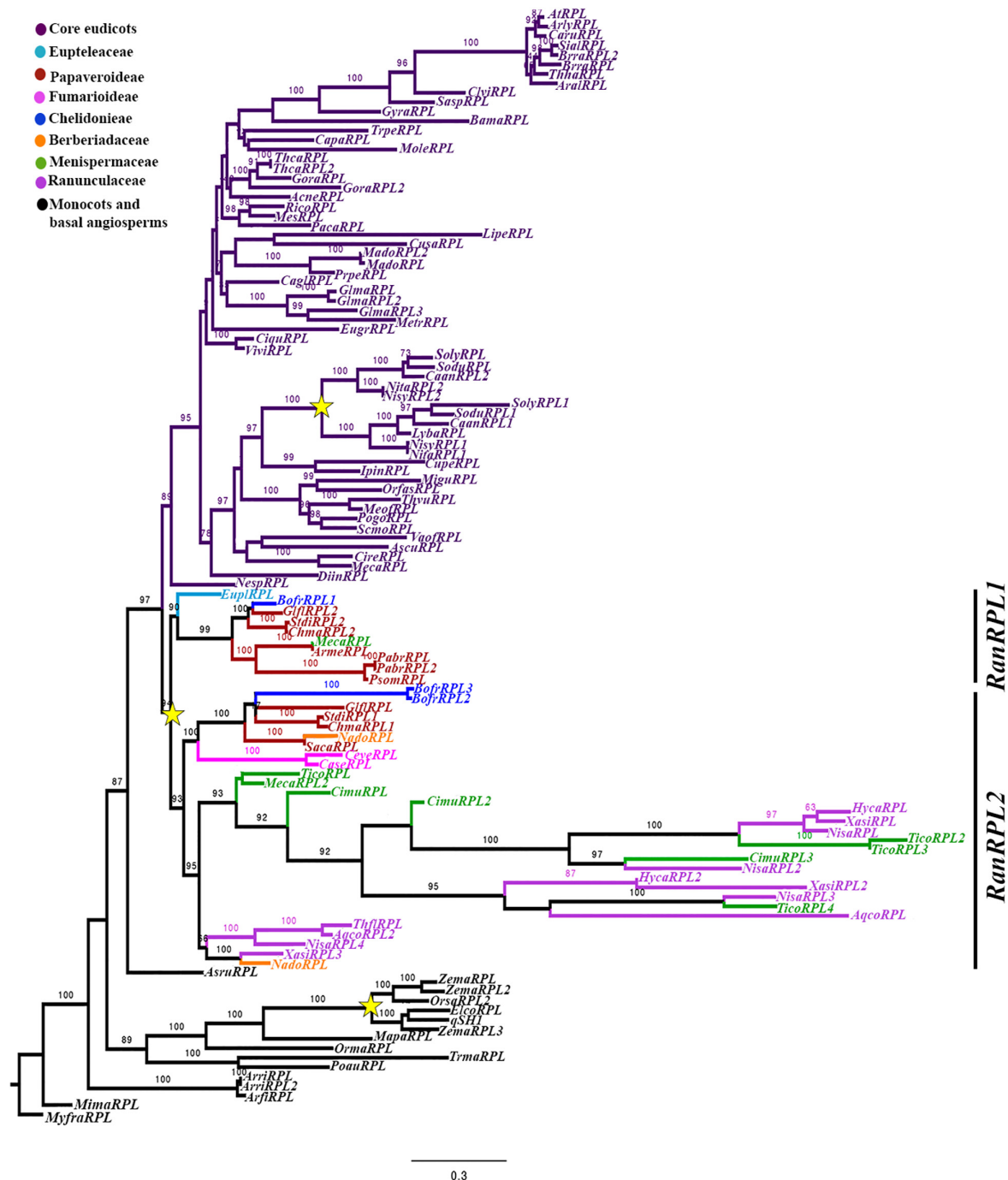


FIGURE 2 | Maximum Likelihood analysis for the *RPL* gene lineage in angiosperms with expanded sampling in basal eudicots using three *PNF* sequences from *A. thaliana*, *A. lyrata* and *C. rubella* (Brassicaceae) as an outgroup. Bootstrap values, larger than 60 are shown at nodes. Two clades were identified for Ranunculales (basal eudicots), *RanRPL1* and *RanRPL2*. Yellow stars indicate two large-scale duplication events, the first one in the Poaceae (Pabón-Mora et al., 2014) and the second one previous to the diversification of the Solanaceae (Ortiz-Ramírez et al., 2018).

CaseRPL; **Figure 3**). Additionally, we identified 14 new motifs that are highly conserved within basal eudicots (**Supplementary Figure S2**). Nine of these motifs (Motifs 9-14, 16, 17, and 20) are lacking in the Ranunculaceae and Menispermaceae protein sequences: HycarPL, NisaRPL1/2/3, XasiRPL2, CimuRPL3,

TicoRPL4, and AqcoRPL, (**Figure 3**). These differences may explain the long branch formed in the *RanRPL2* clade and suggest a relaxed negative selection (**Figure 2**). Interestingly, the MEME analysis found that located before the BELL domain is motif 10, FVDQDC/SCLMESSEDRDCSDDQDEHHHWR



(**Supplementary Figure S2**), which is rich in negatively charged and hydrophobic amino acids and is exclusively found in *Papaver* sequences (**Figure 3**).

***Bocconia frutescens* Flower and Fruit Development**

To better hypothesize the role of *RPL* homologs in basal eudicots, we examined their expression throughout the 11 flower and fruit developmental stages of *Bocconia frutescens* (Papaveraceae) that have been previously defined (Zumajo-Cardona et al., 2017). In addition, we provide anatomical details during later stages of fruit development. Briefly, the determinate inflorescences of *Bocconia frutescens* L., has numerous apetalous flowers that develop basipetally (**Figure 4A**). All flowers are composed of two sepals, a single whorl of homeotic stamens replacing petals, two- three whorls of true stamens and a bicarpellate gynoecium (**Figures 1E, 4A**) (Arango-Ocampo et al., 2016). The two sepal primordia initiate at stage 2, followed by the initiation of the first whorl of homeotic stamens during stage 3 (**Figure 4A**). At stage 4, all the staminal whorls are formed, usually 2 to 3 (**Figure 4A**). The two carpels start developing during stage 5, so that by stage 6 the two carpels overtop the single ovule (**Figure 4A**). During stage 7, the apical region of the carpel differentiates forming the style and stigmas with each carpel composed of 8–10 cell layers. At stage 7, the carpels meet in a commissural ring-like tissue and four main vascular traces are apparent; one in each carpel and one in each of the commissural rings and the cells of the commissural rings are not as densely stained. Also during stage 7, the inner integument of the ovule begins to develop (**Figure 4B**). During floral developmental stages 8–9, three main proximo-distal zones differentiate: the gynophore, the ovary and the style (**Figures 4C,D**). The style rapidly extends and is topped with two massive papillate stigmas (**Figure 4D**). At stage 8, the ovary wall is composed of 12 cell layers and a 4–5 cell separation layer, between the carpels and the commissural ring, becomes distinct (**Figure 4C**). Also at stage 8, the ovule is developing with the inner integument covering the nucellus while the outer integument is still elongating (**Figure 4E**). At stage 9 (**Figures 4D,E,G**), the two integuments completely cover the single anatropous ovule, forming what is going to be the seed coat; the ovary walls are formed by 12–15 cell layers of parenchymatous tissue (**Figures 4D,E,G**), with three vascular bundles (one central and two lateral) vascular bundles in each carpel and a vascular bundle in the commissural ring, as in stage 8 (**Figure 4E**). As detailed fruit developmental stages have not been described, here, we provide cross sections of the fruits at distinct stages (**Figures 4H–J**). After pollination, the two carpels become the two valves. The number of vascular bundles in the valves and the commissural ring remains the same (**Figure 4H**). During the first stage of fruit development, stage 10 (**Figure 4H**), the 12–15 cells forming the pericarp become more compressed and can be further distinguished into the exocarp with a single cell layer of rectangular tightly packed cells, covered by a thin cuticle, the mesocarp formed by larger isodiametric parenchymatous cells with three vascular bundles, and the endocarp formed by smaller cells than the exocarp. The separation layers between

the valves and the persistent commissural ring are composed of 2–3 cell layers of densely stained cells. The commissural ring is wedge-shaped, formed by parenchymatous tissue with one vascular bundle (**Figure 4H**). Also during this stage the seed is fed by a vascular bundle and it is possible to differentiate the endosperm and the seed coat which develops from the two integuments, the three inner layers of the seed coat arise from the inner integument and the 6–8 outer layers from the outer integument (**Figures 4E,G,I**). During stage 11 of fruit development, (**Figures 4I,J**), the pericarp becomes differentiated into an exocarp of radially elongated and enlarged cells, a mesocarp formed by 6–7 of tangentially elongated cells and, an endocarp with 3–4 layers of very small flat cells (**Figure 4I**). The epidermal cells of the commissural ring are different from the exocarp of the valves; cells in the commissural ring are smaller (**Figure 4I**). When the fruit ripens as stage 12 the two valves fall off leaving only the seed attached to the commissural ring (see *Bocconia* fruit, **Figure 1E**). The base of the seed is covered with a red aril that originates from the funiculus after fertilization; during the ripening of the seed it does not form part of the seed coat as it develops a distinctive coloring and texture (**Figures 4I,J**). The aril in *Bocconia* serves to attract birds since the brilliant red color of it contrasts with the black testa (see *Bocconia* fruit, **Figure 1E**). Arils occur on seeds of many tropical species. Interestingly, *Bocconia* is the only neo-tropical genus from the Papaveraceae and the only one within the family that has an aril.

Expression of *RPL Bocconia frutescens* Homologs (*BofrRPL1/2/3*)

We evaluated the expression patterns of the three *REPLUMLESS* paralogs (*BofrRPL1/2/3*) in *B. frutescens* with specific probes designed for each one (**Supplementary Table S2** and **Supplementary Figure S1**). Our results show different expression patterns for the three *Bocconia frutescens* *RPL* homologs. *BofrRPL1* expression was not detected in the vegetative meristem, young leaves (**Figure 5A**) or during the initiation of floral organ primordia at stages 3–6 (**Figures 5B–E**). Low levels of *BofrRPL1* expression are detected at the sepal tips, specifically in the vascular traces during stages 4–5 (**Figures 5D,E**). Expression of *BofrRPL1* is stronger later in flower development, at stage 6 when the two carpels overtop the single ovule, where *BofrRPL1* is detected in the sepal tips, stamens, both in the anthers and the filaments, the adaxial side of the carpels and in the tip of the nucellus during ovule elongation prior to integument initiation (**Figure 5F**). During stages 7 and 8 when differentiation of the style and stigma occur, *BofrRPL1* expression is maintained in the sepal tips, the stamens, the adaxial region of each style and stigma toward the medial region where fusion will occur, and the ovule (**Figures 5E,G**). However, in the fully differentiated gynoecium, after syncarpy, *BofrRPL1* becomes restricted to the proximal region of the long styles specifically toward their adaxial epidermis where fusion has occurred (**Figure 5H**). At this stage, expression of *BofrRPL1* was also found in the vasculature at the base of the flower (**Figures 5G,H**). During the transition to fruit development, between stages 9–10, *BofrRPL1* is expressed in the 3–4 cell layers

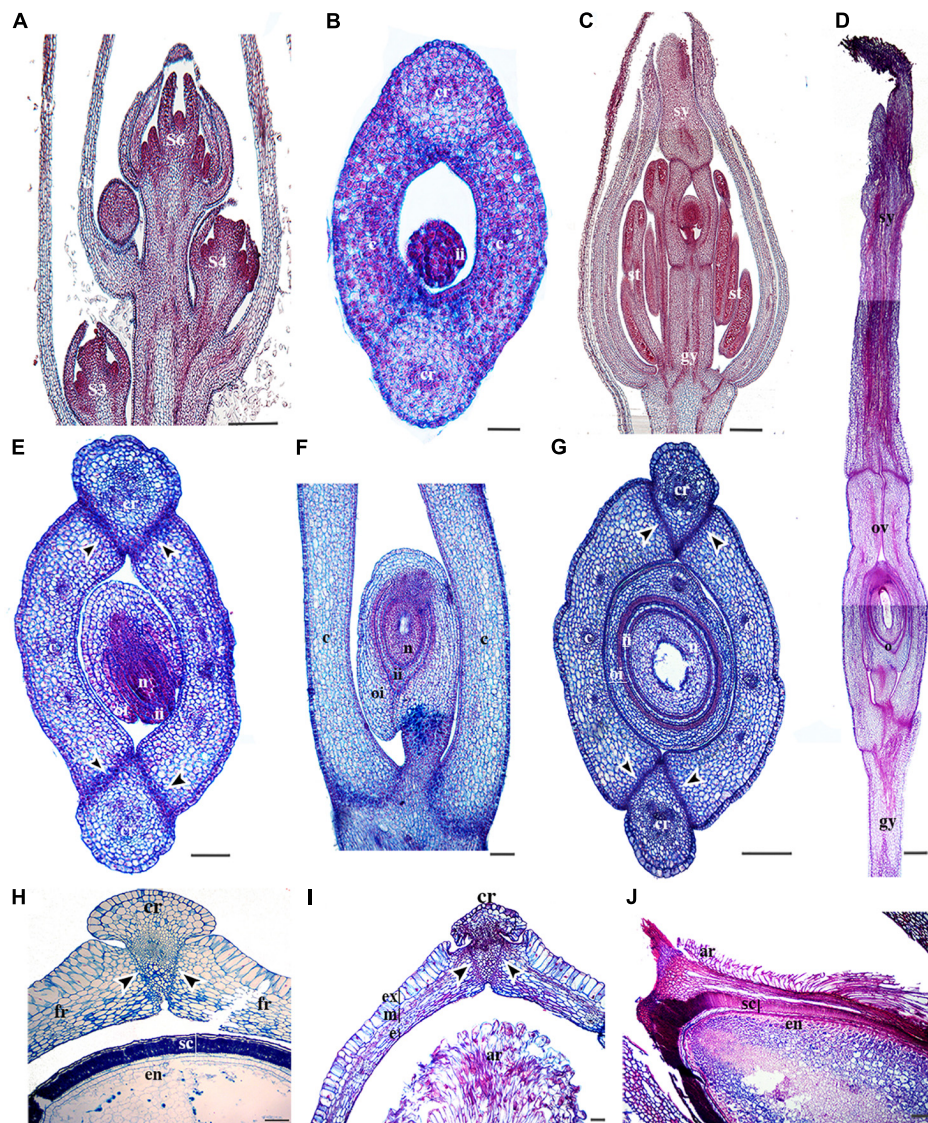


FIGURE 4 | Anatomy of *B. frutescens* flowers and fruits. **(A)** Longitudinal section of an inflorescence with flowers at different stages (S): S3, S4, and S6. Stages follow Zumajo-Cardona et al. (2017). **(B)** Cross-section of a carpel at S6. **(C)** Longitudinal section of a flower at S8. **(D)** Longitudinal section of the gynoecium at S9, showing the three main proximo-distal zones: gynophore, ovary, and style with the two stigmas at the apex. **(E)** Cross-section of a flower at S8. **(F)** Close-up longitudinal section of an ovary at S9. **(G)** Cross-section of the ovary at S9. **(H)** Cross-section of a young fruit at S10 and the seed. **(I)** Cross-section of a fruit at S11 showing fruit layers: exocarp, mesocarp and endocarp. **(J)** Longitudinal section of the seed coat and aril. Black arrowheads indicate the separation layer, ar, aril; b, bract; c, carpel; cr, commissural ring (persistent medial tissue); e, endocarp; en, endosperm; ex, exocarp; fr, fruit wall; gy, gynophore; ii, inner integument; m, mesocarp; n, nucleus; o, ovule; oi, outer integument; ov, ovary; s, sepal; sc, seed coat; se, seed; st, stamen; sy, style; v, valve. Scale bars: 100 μm (**B,I-J**), 200 μm (**D**), 250 μm (**H**), 0.2 mm (**E**), 0.5 mm (**A,C,F,G**).

between the valves and the commissural ring that will form the dehiscence zone in the mature fruits (**Figures 5I,J**). *BofrRPL1* is also expressed in the vascular bundle that feeds the commissural ring as well as in the three vascular bundles found in each valve (**Figures 5I,J**).

Unlike *BofrRPL1*, *BofrRPL2* is detected in early vegetative and floral development. *BofrRPL2* is found during vegetative development on the adaxial side of the leaf primordia where it is maintained during leaf growth (**Figure 6A**). During early flower development, between stages 3–6, *BofrRPL2*

is expressed in the sepals, particularly at their tips, in stamen and carpel primordia (**Figure 6B**), as well as in the ovule primordium (**Figures 6C–E**). During stages 7–9, the expression of *BofrRPL2* decreases dramatically, nevertheless it is still detected in the sepal tips, the anthers, the adaxial side of the carpels where the styles will fuse and the tips of the ovule where the two integuments will develop (**Figures 6F–H**). During fruit development at stages 10–11, *BofrRPL2* expression is found in cell layers between the valves and the commissural ring, the region that will correspond to

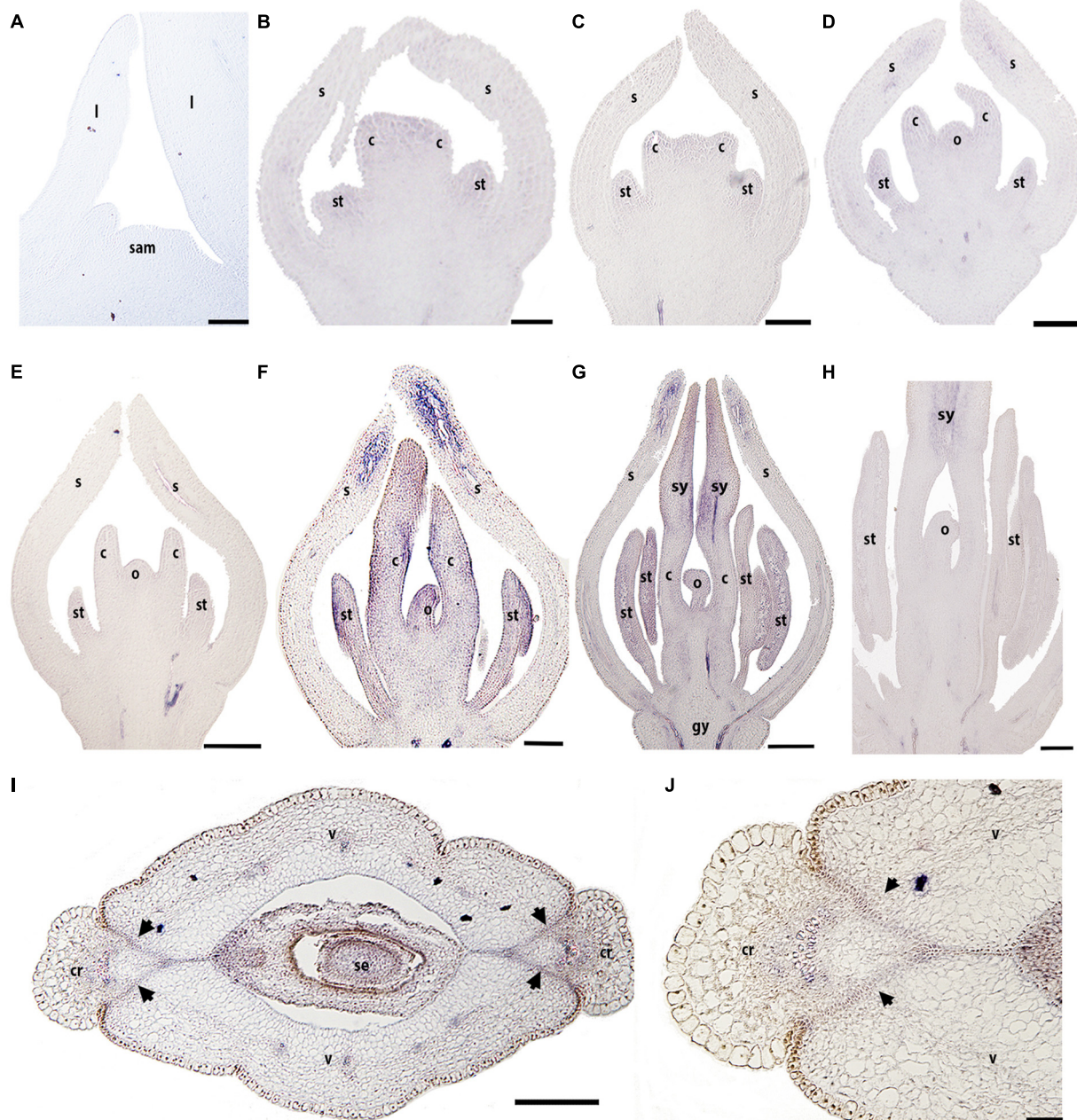


FIGURE 5 | Expression of *BofrRPL1* by *in situ* hybridization in longitudinal (A–H) and cross-sections (I, J) of developing shoots, flowers, and fruits. (A) *BofrRPL1* expression is not detected in the shoot apical meristem, (B) floral bud in stage 3, (C–E) or floral stages 4–6. (F) *BofrRPL1* expression is first detected at floral stage 7 in stamens, carpels, and ovule. (G, H) At stage 8 *BofrRPL1* is restricted to the region where the two stigmas fuse floral. (I, J) *BofrRPL1* expression becomes restricted during fruit development and is detected in the dehiscence zone. Black arrows indicate the dehiscence zones, b, bract; c, carpel; cr, commissural ring; gy, gynophore; l, leaf; o, ovule; s, sepal; sam, shoot apical meristem; se, seed; st, stamen; sy, style; v, valve. Scale bars: 50 μ m (A–C) 100 μ m (D), 0.1 mm (E–H), 0.2 mm (I), 500 μ m (J).

the dehiscence zone during fruit ripening, as well as in the seed (Figures 6I, J).

The expression of *BofrRPL3* is similar to *BofrRPL2* in regard to its early expression in vegetative and floral development. *BofrRPL3* expression is detected during vegetative development in the shoot apical meristem, in the leaf primordia as well as in

the procambium and the vascular traces feeding the young leaves. It is also expressed in the adaxial region of more mature leaves (Figure 7A). The expression of *BofrRPL3* during early floral development (stages 3–6) is more similar to the expression found for *BofrRPL2* than to *BofrRPL1*, as it is strongly expressed in the sepal tips, the stamen and carpel primordia and during ovule

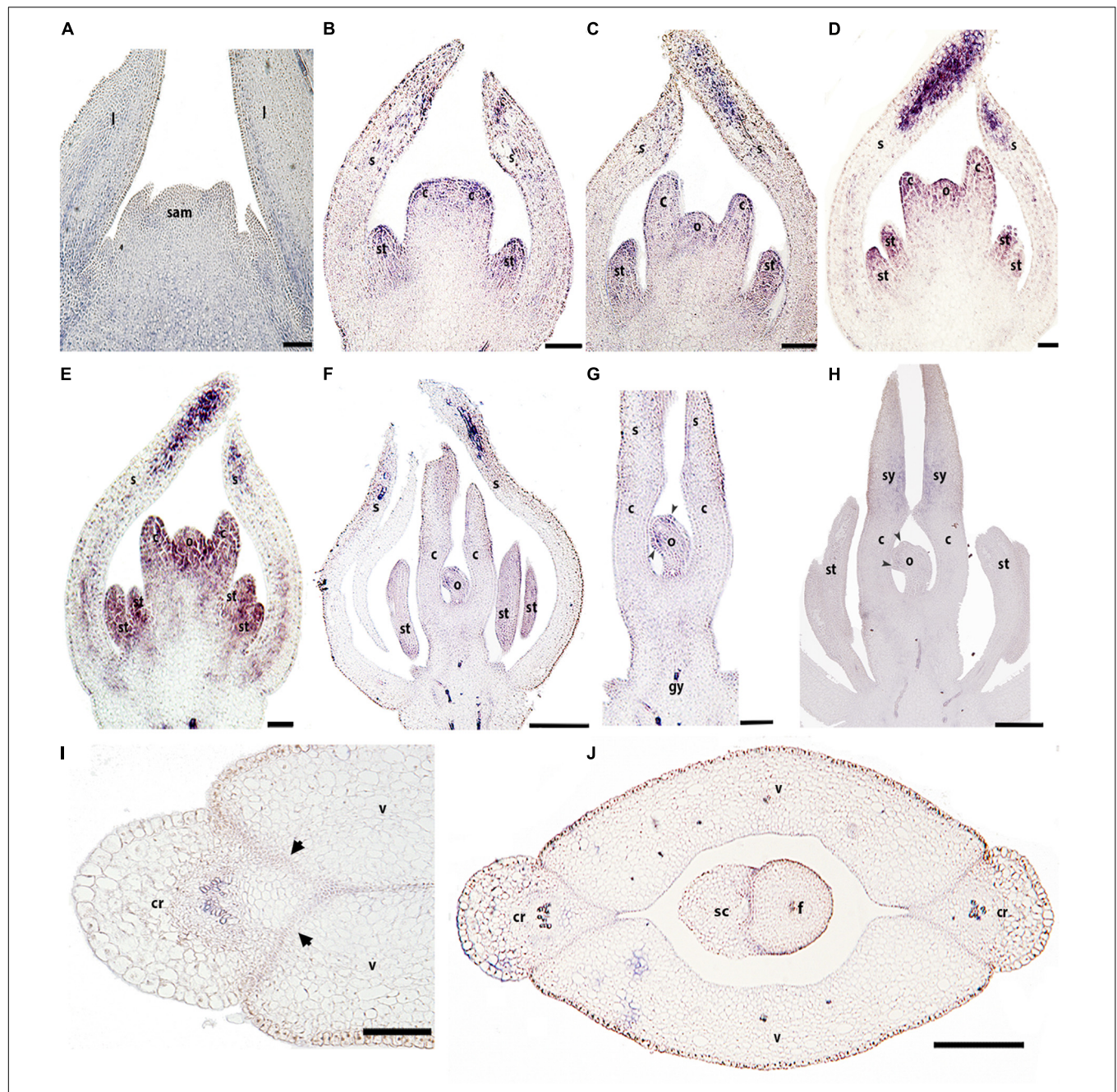


FIGURE 6 | Expression analyses of *BofrRPL2* by *in situ* hybridization in longitudinal (A–H) and cross-sections (I, J) of developing shoots, flowers, and fruits. (A) *BofrRPL2* expression is detected in the adaxial side of leaf primordia and in more mature leaves. (B–F) *BofrRPL2* expression is detected in floral stages 3–7 in stamen and carpel primordia. (C–E) These expression patterns are maintained from stages 4 to 6, where *BofrRPL2* is also expressed in the ovule primordia. (F) At stage 7 the expression in the ovule is restricted to the initiation of the integuments. (G) *BofrRPL2* expression is only detected in the style and integument primordia during floral stages 8–9. (I–J) *BofrRPL2* expression is restricted to the dehiscence zone during fruit development. Black arrowheads indicate integument primordia, Black arrows indicate the dehiscence zones of the fruit, b, bract; c, carpel; cr, commissural ring; f, funiculus; gy, gynophore; l, leaf; o, ovule; s, sepal; sam, shoot apical meristem; sc, seed coat; se, seed; st, stamen; sy, style; v, valve. Scale bars: 50 μ m (A), 100 μ m (B–E), 0.1 mm (H), 0.2 mm (F, G, I, J).

initiation (Figures 7B–D). Later during flower development, at stage 7, the expression of *BofrRPL3* is strongly maintained in the stamens and in the growing tips of the two carpels that fuse to each other enclosing the ovule (Figure 7E). During carpel development at stages 8–9, *BofrRPL3* is found in the sporogenous

tissue of the anthers, toward the adaxial surface of each elongating style, at the tip of the ovule and in the vasculature of the receptacle (Figures 7F–H). *BofrRPL3* is expressed throughout ovule development (Figures 7D–H). During fruit development at stage 10, *BofrRPL3* expression is detected in the 3–4 cell layers

of the separation layer between the valves and the commissural ring as well as in the carpel wall (Figures 7I,J). However, this expression is not maintained in the mature fruits at stage 11, where *BofrRPL3* is only restricted to the aril (Figure 7K). In fact, it is the only paralog showing this expression pattern and likely reflects neofunctionalization (Figures 5–7).

Papaver somniferum Carpel and Fruit Development

Our descriptions of the expression analyses for *PsomRPL* in *Papaver somniferum* follow those of Drea et al. (2007) where stages P3, P5, P7, and P8 were defined (hereafter referred to without the P) and Becker et al. (2005), but include a number of previously undescribed developmental stages. We defined floral and fruit developmental stages based on the following developmental landmarks (Supplementary Table S3 and Figure 8). At stage 1, the floral meristem can be distinguished and the sepal primordia are initiating (Figure 8A). Stage 2 is defined by petal initiation. By stage 3 (Drea et al., 2007), stamen and carpel primordia initiate, and the sepals enclose the floral bud (Figure 8B). Stage 4 is defined by the elongation of the multiple carpels. At stage 5 (Drea et al., 2007), the filament of the stamen is distinguishable from the anther and the multiple carpels forming the gynoecium are well differentiated (Figure 8C). At stage 6, the carpels start to fuse (Figure 8D). During stage 7 (Drea et al., 2007), the petals grow and fill the space bounded by the two sepals and in the multicarpellate gynoecium, the ovules initiate on the carpel walls to give parietal placentation, and the stigmatic lobes begin to differentiate (Figure 8E). During stage 8 (Drea et al., 2007), the pedicel of the preanthetic flower undergoes asymmetric growth; the floral buds become pendant right before anthesis and then extend upright when the flower opens in the subsequent stage. Also during stage 8, the fully fused stigmatic tips develop the characteristic upright papillae crowning the distal portion of the gynoecium. The carpel wall is formed by a 11–12 cell layer at this stage (Figure 8F) and remains the same at anthesis corresponding to stage 9 (Figure 8G). Stage 10 is defined by post-fertilization development; the lobes of the stigmas fuse to each other at the tip as a result of residual meristematic activity (Figure 8H), forming a crown-like stigmatic ring (Figure 8I). The fruit wall is formed of 14–18 cell layers, where the exocarp, mesocarp and endocarp are formed of parenchyma cells with small intercellular spaces (Figure 8I). There are multiple vascular bundles in the fruit wall but each carpel has a massive vascular bundle that can be distinguished at the position where the placenta is formed (Figure 8I). At stage 11, the fruit is mature and the fruit wall becomes more compressed with 12–14 cell layers apparent (Figures 8J–L). The apical stigmatic ring is formed of 6–8 cell layers and later during fruit maturation each of the stigmatic rays will produce tension outward antagonizing the poricidal dehiscence zones in each locule (Figures 8K,L). In the poppy variety used in this study, *P. somniferum* cv. Persian White, the fruit does not open leaving the seeds enclosed. Therefore, we will refer here to the dehiscence layer between the fruit wall and the persistent stigmatic ring as a putative separation layer. The

putative separation layer is formed of approximately two cell layers between the stigmatic ring and the fruit wall in each of the locules (Figure 8K). Laticifers can be found throughout the pericarp (Figures 8H–L). Fruit dehiscence in wild *P. somniferum* occurs by pores (stage 12), where tension occurs apically below the stigmatic ring and in between the central vascular bundles holding the parietal placenta and the fruit wall in each of the locules leaving only apical pores (Figures 1A,B) (Roth, 1977). The number of pores formed in the pericarp corresponds to the number of locules in the fruit. When the fruit is completely lignified, the rays of stigmas bend outward, these separate along the separation layer between two vascular carpellary central bundles.

Expression of a RPL Homolog in Papaver somniferum

To better understand the role of *RPL* in Papaveraceae, we analyzed the expression of the single *RPL* homolog identified in *Papaver somniferum* (*PsomRPL*). *PsomRPL* expression is detected during vegetative development in the stem as well as in the shoot apical meristem and the adaxial region of the emerging leaves (Figure 9A). The expression in the floral vascular bundles is maintained throughout floral development between stages 1–9 (Figures 9B–I). *PsomRPL* expression is first detected in the flower at stage 3, where it is found at the tip of the sepals enclosing the floral bud (Figures 9C,D). During stage 5, *PsomRPL* is expressed in the petal primordia, the stamens as well as in the growing tips of the carpel primordia (Figure 9E). At stage 6, *PsomRPL* is expressed in between the floral organs where their proximal portions connect with the receptacle, as well as in the growing petals, stamens and the carpels (Figure 9F). Later during stage 7, the expression in the stamens is restricted to the filament (Figure 9G) as well as to the carpel wall. *PsomRPL* expression is also detected at the junction of each floral organ on the floral receptacle and is maintained during stages 7 and 8 (Figures 9G,H). At stage 8 *PsomRPL* is also expressed in the sporogenous tissue of the anthers and the developing ovules (Figure 9H). At stage 9, *PsomRPL* is differentially expressed in the carpel; it is detected in the region where the carpels fuse, in the extending parietal placentas, and in the endocarp and mesocarp (Figures 9I,J). Later, in the young fruit, *PsomRPL* is expressed in the cells that constitute a putative separation layer between the fruit wall and the stigmas (Figure 9K), in the vascular bundles, the placenta, the epidermis of the fruit wall, and the laticifers (Figures 9L–N).

DISCUSSION

Very little is known about the fruit developmental network outside the Brassicaceae. *RPL*, particularly, has been described for its function in the proper development of the replum, tissue that is only found in the Brassicaceae fruits (Alvarez and Smyth, 2002; Roeder et al., 2003). Although *RPL* does not specify replum identity directly, it is a repressor of valve margin and valve identity as specified by *SHATTERPROOF*, *INDEHISCENT* and

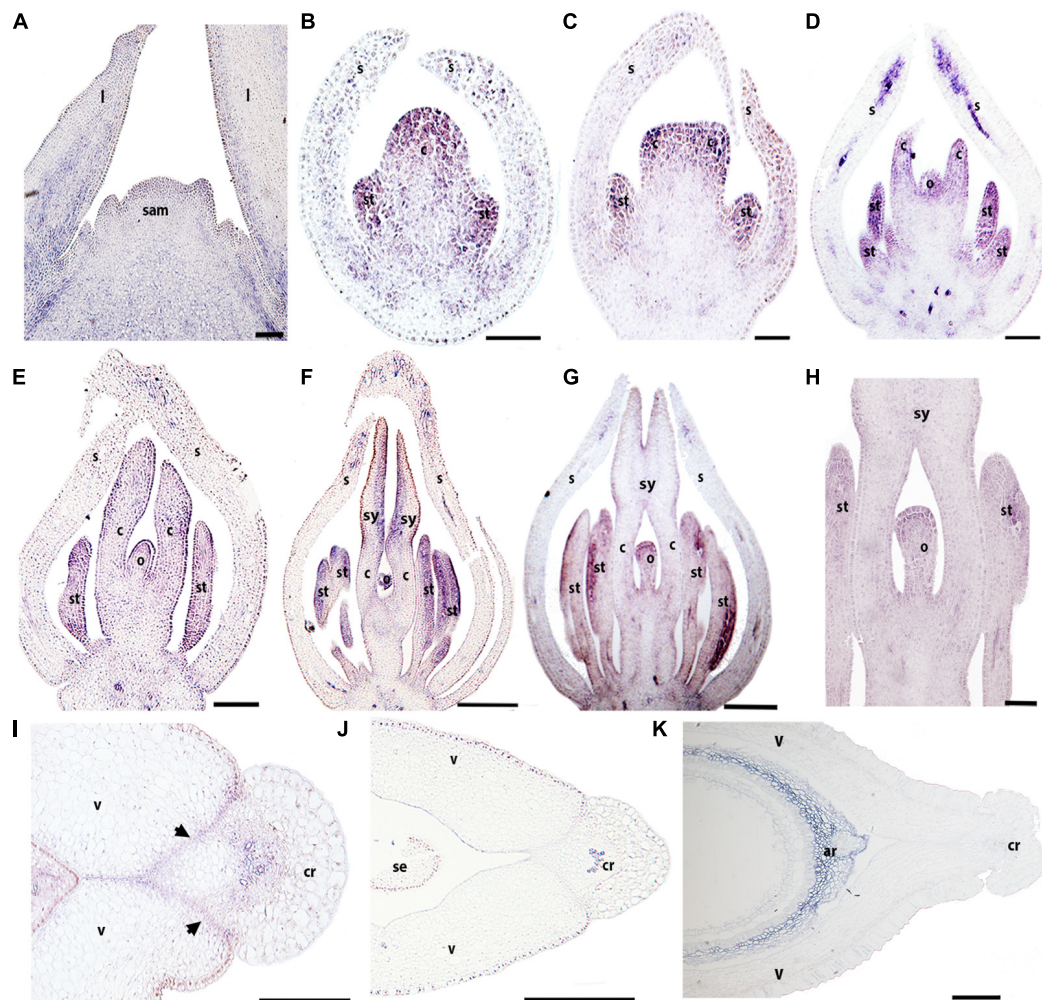


FIGURE 7 | Expression of *BofrRPL3* by *in situ* hybridization of longitudinal (A–H) and cross-sections (I–K) of developing flowers and fruits. (A) *BofrRPL3* expression is first detected in the apex of the shoot apical meristem, in leaf primordia and in the adaxial side of more developed leaves. (B–F) *BofrRPL3* expression is detected in floral stages 2–7. (B) *BofrRPL3* expression is detected in the stamen and carpel primordia as they emerge and (C) as the two carpel primordia begin to elongate. (D–H) *BofrRPL3* expression persists in stamens, in carpels as they elongate and is also detected throughout ovule development up until floral stages 8–9. (I) *BofrRPL3* expression is detected in the gynoecium at S9 between the valves and the commissural ring. (J) *BofrRPL3* expression becomes restricted to the epidermis during fruit development. (K) *BofrRPL3* expression is restricted to the aril during seed development. Black arrows indicate the dehiscence zones. ar, aril; b, bract; c, carpel; cr, commissural ring; l, leaf; o, ovule; sam, shoot apical meristem; s, sepal; se, seed; st, stamen; sy, style; v, valve. Scale bars: 50 μ m (A,K), 100 μ m (B–D), 0.1 mm (F–I), 0.2 mm (E,J).

FRUITFULL (Roeder et al., 2003; Girin et al., 2009). In addition, *RPL* has pleiotropic roles in *Arabidopsis* development including meristem identity, inflorescence, and fruit development as indicated by its many synonyms: *BELLRINGER*, *PENNYWISE*, and *VAAMANA* to name a few (Byrne et al., 2003; Smith and Hake, 2003; Bhatt et al., 2004; Hake et al., 2004). The function of *RPL* orthologs in other angiosperms is only known from the monocot crop species, rice. In rice, it is involved in fruit shedding, and therefore is one of the genes involved in its domestication (Konishi et al., 2006; Arnaud et al., 2011). More recently, expression analyses in selected species of *Solanaceae* have shown similar expression patterns during flower and fruit development as in *Arabidopsis* suggesting conserved roles (Ortiz-Ramírez et al., 2018). Expression and functional data

available for *RPL* genes are not sufficient to assess their functional evolution across angiosperms. Due to limited functional studies in non-model, non-core eudicots, expression analyses provide a solid base to better predict the functional evolution of this gene lineage. Here, we discuss the expression patterns of *RPL* homologs in two basal eudicots, *Bocconia frutescens* and *Papaver somniferum* (Papaveraceae) whose phylogenetic position is crucial to understanding the functional differences observed between rice (monocots) and core eudicot *RPL* genes; additionally, the two species selected in this study present dry dehiscent fruits with different mechanisms for seed dispersal allowing comparative analyses of *RPL* contribution to opercular and poricidal dehiscence (Kadereit, 1993).

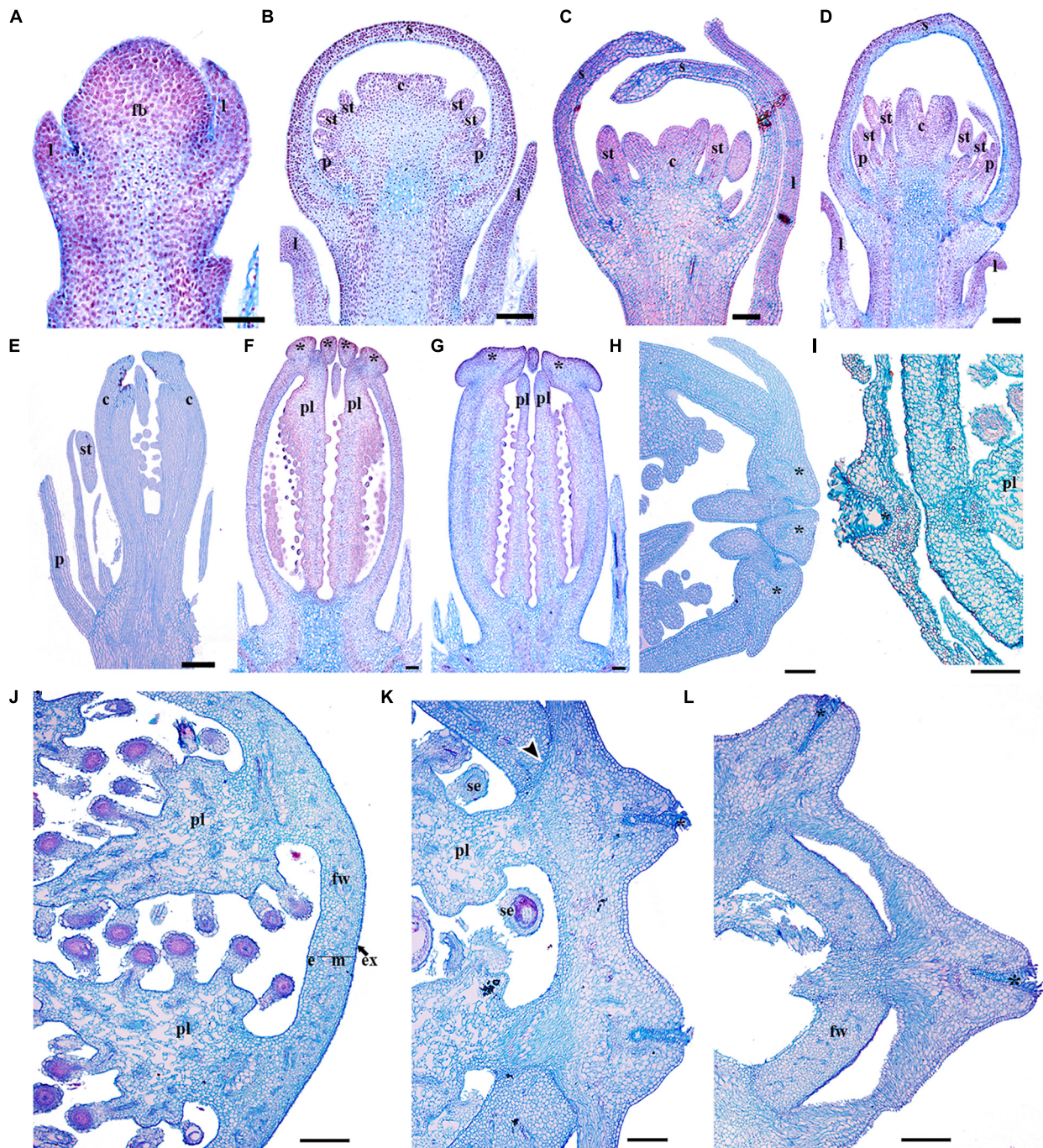


FIGURE 8 | Flower and fruit developmental stages of *Papaver somniferum*. **(A)** Floral bud in stage 1 during sepal initiation. **(B)** Floral bud in stage 3 when petals, stamens and carpel primordia can be distinguished. **(C)** Floral bud in stage 5 with the initiation of a multicarpellate gynoecium and the filament of the stamens. **(D)** Floral buds in stage 6 with the carpels overtopping the multiple ovules. **(E)** Flowers at stage 7 when the initiation of the stigmatic region occurs and the ovules develop by parietal placentation are clearly distinguished. **(F)** Gynoecium in pre-anthesis, stage 8. **(G)** Carpel of a flower in anthesis, the lobules of the stigmas start to elongate. **(H)** Apical region of a young fruit. The tip of the fruit is shown to the right. **(I)** Cross-section of a young fruit showing the apical region with papillose stigma and the fruit wall. **(J)** Cross section through the mid region of a more mature fruit **(K)** Close-up of the crowning stigmatic ring, showing the putative separation layer between the styles and the fruit wall. **(L)** Cross section across the apex of the fruit showing the crowning stigmatic ring. Asterisks indicate stigmas, Black arrowheads indicate the putative dehiscence region of the fruit, c, carpel; e, endocarp; fb, floral bud; fw, fruit wall; m, mesocarp; p, petal; pl, placenta; s, sepal; se, seed; st, stamen. Scale bars: 50 μ m **(A–E)**, 250 μ m **(F–H, J–L)**, 500 μ m **(I)**.

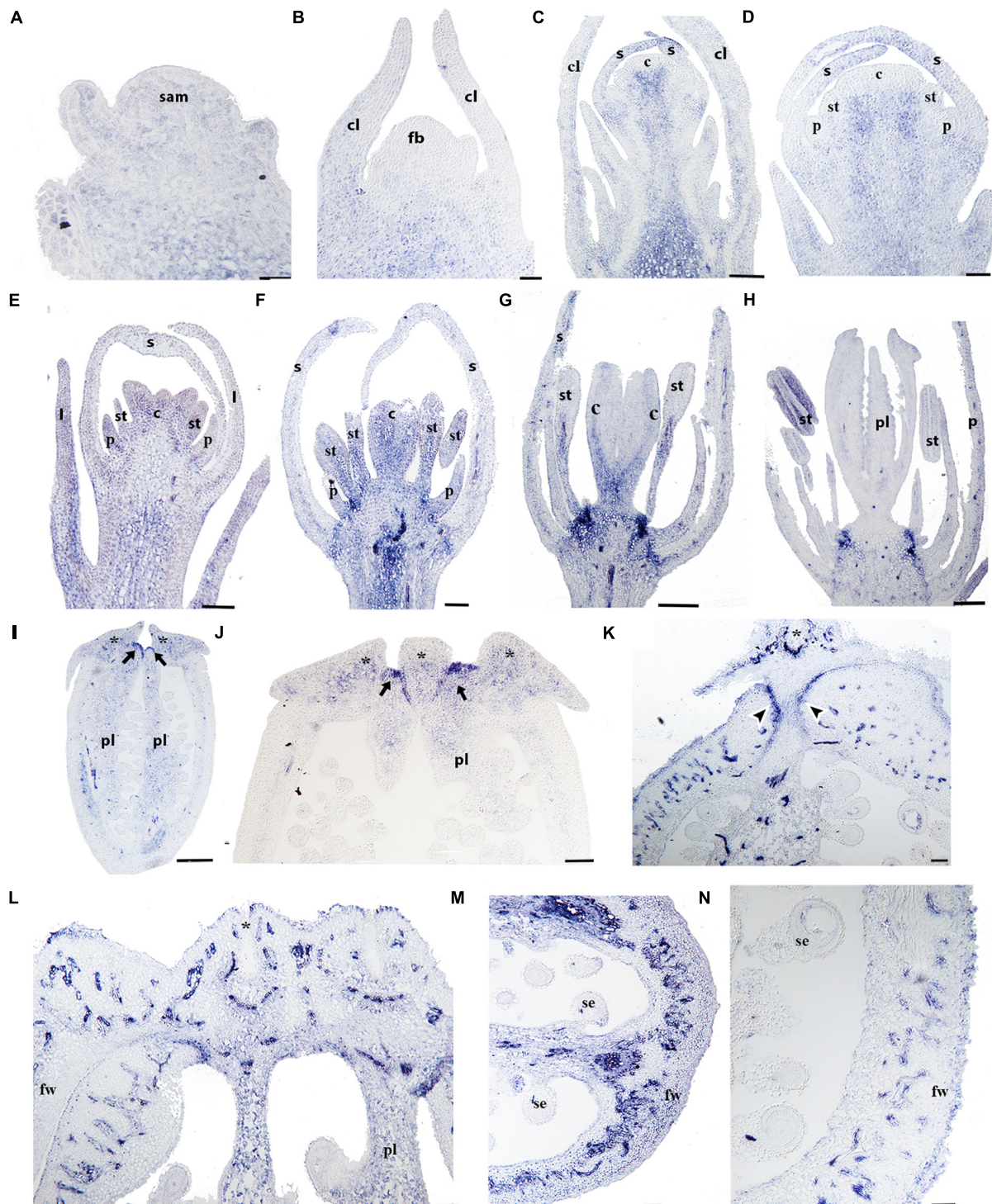


FIGURE 9 | Expression analyses of *PsomRPL* by *in situ* hybridization. Longitudinal (A–L, N) and cross-sections (L–N) of developing shoots, flowers, and fruits. (A) *PsomRPL* expression is detected in the shoot apical meristem, the adaxial region of the leaf primordia and the stem. (B) No expression is detected in the floral bud at stage 1. (C, D) During stage 3, *PsomRPL* is expressed in the sepal tips. (E–G) At floral stage 5–7, *PsomRPL* expression is detected in the petal, stamen, and carpel primordia. (H) During pre-anthesis expression becomes restricted to the ovules and on the receptacle in between the fusion of the floral organs. (I, J) At stages 9–10, *RPL* is expressed in the style and where the stigmas will form and in the placenta. (K–N) In young fruits, *PsomRPL* expression is detected in the vascular bundles, the putative separation layer and the apical region where all the carpels fuse. Asterisk stigma, Black arrows point to the apical region where the carpels fuse, Black arrowheads point to the putative separation layer of the fruit, c, carpel; cl, cauline leaf; fb, floral bud; fw, fruit wall; p, petal; pl, placenta; s, sepal; se, seed; st, stamen. Scale bars: 50 μ m (B–F, K, L–N), 100 μ m (A, G), 200 μ m (H), 500 μ m (I), 0.2 mm (J).

There Are Two *RPL* Clades Within Ranunculales: *RanRPL1* and *RanRPL2*

According to previous studies *RPL* genes have evolved with the radiation of angiosperms as the result of a duplication event predating angiosperm diversification resulting in the *RPL* clade and its sister clade *POUNDFOOLISH* (*PNF*; Pabón-Mora et al., 2014) (Figure 2). *RPL* are predominantly single copy genes in angiosperms with few exceptions, particularly in taxa corresponding to recent polyploids, such as *Bocconia frutescens*, *Glycine max*, *Malus domestica*, *Papaver bracteatum*, *Theobroma cacao*, and *Tinospora cordifolia* (Figure 2) and in members of the Solanaceae and Poales, where independent duplications have been found coinciding with known ancient WGD events (Jiao et al., 2011; D'Hont et al., 2012; Pabón-Mora et al., 2014; Ortiz-Ramírez et al., 2018). We performed an exhaustive search in publicly available databases for *RPL* homologs from basal eudicots and we were able to retrieve sequences from all families within Ranunculales. Here, we report an additional duplication, previously unidentified, likely predating the diversification of Ranunculales, resulting in two clades: *RanRPL1* and *RanRPL2* (BS = 94). However, only one gene of *Euptelea pleiosperma* was found nested in the *RanRPL1* clade so it is likely that the duplication event occurred before the radiation of Eupteleaceae but the paralog in the *RanRPL2* clade has not yet been identified (Figure 2). In addition, the two clades vary in terms of the plant groups with representative *RPL* sequences in each. *RanRPL1* only includes sequences of Eupteleaceae, Menispermaceae, and Papaveraceae while *RanRPL2* includes Berberidaceae, Menispermaceae, Papaveraceae, and Ranunculaceae sequences (Figure 2). These differences indicate that some *RPL* sequences were not retrieved even after extensive searches in available transcriptomes (for instance in the case of *Eschscholzia californica*). This can be due to (1) *RPL* gene expression in tissues or organs different from which the transcriptomes were generated in a case by case scenario, (2) low expression of *RPL* copies resulting in low depth and coverage of contigs that remain undetected in the assemblies available (20M), or (3) true gene losses that are harder to assess until more genomes become available. In fact, although we were able to identify two *RPL* copies in *Aquilegia coerulea* (Ranunculaceae), the only basal eudicot with the genome sequenced, no Ranunculaceae sequences were retrieved for the *RanRPL1* clade suggesting that this clade has been lost at least in Ranunculaceae. In addition, these two clades differ in substitution rates and this becomes evident with the long branch formed in the phylogenetic analysis which includes most of the Ranunculaceae and Menispermaceae sequences (BS = 92; Figure 2).

An examination of the conserved domains across basal eudicots (Figure 3), showed the previously identified domains in *RPL* homologs: the Homeodomain near the C-terminus, (Figure 3B), and the MEINOX INTERACTING DOMAIN (MID) near the N terminus. The MID domain is composed of the SKY and BELL-domains (Hake et al., 2004; Hay and Tsiantis, 2009; Mukherjee et al., 2009). We found that the BELL-domain is conserved in basal eudicots but the SKY motif can be replaced sometimes by SRF (Motif 5,

Figures 3A,B) similar to previous results (Pabón-Mora et al., 2014). The location of the ZIBEL motif has been previously found to occur both before the BELL domain and after the Homeodomain (Mukherjee et al., 2009). For the basal eudicot sequences analyzed here, we were able to find the ZIBEL motif (Motif 6 according to the MEME analysis) in some proteins before the BELL domain and in others toward the N terminal region of the protein but never in both locations (Figure 3A) (Pabón-Mora et al., 2014). The MEME analysis allowed us to also identify additional conserved motifs that have not been functionally characterized (Supplementary Figure S2). For example, motif 7 is highly conserved in basal eudicot *RPL*s and it is located between the BELL and Homeodomain (Figure 3 and Supplementary Figure S2) region, which has been described as highly variable (Mukherjee et al., 2009). Motif 10 is rich in hydrophobic and negatively charged amino-acids and is present only in the *Papaver* proteins included in the analysis: *PsomRPL*, *PabrRPL*, and *PabrRPL2* (Figure 3A) which may impact protein folding and binding (Nicholls et al., 1991) and therefore confer a specific function to these proteins. In addition, motifs 8, 13, and 20 are found only in the sequences belonging to the *RanRPL2* clade but absent in some Menispermaceae and Ranunculaceae proteins such as *HycarRPL*, *NisarRPL1/2/3*, *XasiRPL1*, *TicoRPL4* and *AqcoRPL* (Figure 3A and Supplementary Figure S2). In addition, these proteins are also highly variable toward the N-terminus (Figure 3A).

RPL Expression During Vegetative Development Is Conserved Across Eudicots

To fill the gaps in our understanding of *RPL* evolution across angiosperms and to propose hypotheses in terms of the functional evolution of the *RPL* gene lineage, we analyzed *RPL* expression patterns in *Papaver somniferum* and *Bocconia frutescens* (Papaveraceae; basal eudicots). The two paralogs from *Bocconia* that belong to the *RanRPL2* clade, *BofruRPL2/3*, are expressed in the shoot apical meristem, the adaxial side of the developing leaves as well as in the adaxial side of more developed leaves. The fact that *BofruRPL1* is not expressed in the vegetative tissue suggest some degree of subfunctionalization among the three *B. frutescens* paralogs (Figures 5–7). On the other hand, *PsomRPL*, part of the *RanRPL1* clade, is found to be expressed in the vegetative tissue in shoot apical meristem, developing leaves and in the stem (Figure 9).

RPL homologs analyzed here show expression patterns in the vegetative tissue similar to those found in *Arabidopsis RPL* and its sister clade *PNF* (Becker et al., 2002; Roeder et al., 2003; Konishi et al., 2006; Campbell et al., 2008; Arnaud et al., 2011; Pabón-Mora et al., 2014; Bencivenga et al., 2016). Our results suggest that *RPL* function in the maintenance of shoot apical meristem identity, is conserved in monocots and eudicots (Bhatt et al., 2004; Konishi et al., 2006; Kanrar et al., 2008; Rutjens et al., 2009; Arnaud et al., 2011; Mühlhausen et al., 2013). In *Arabidopsis*, this function is mediated by the interaction with *KNOX* genes (Cole et al., 2006) which also have been found to have a conserved function in Papaveraceae compared to *Arabidopsis* (Groot et al.,

2005). We hypothesize that the meristem identity function mediated by the interaction of *RPL* and *KNOX* genes is likely maintained in eudicots. In fact, its meristematic function may be also maintained in fruit development across eudicots as we found *RPL* expression in *P. somniferum* at the apex of the carpel walls where they fuse as result of the post-genital meristematic activity (**Figures 8, 9**) (reviewed by Girin et al., 2009).

Our results together with those found in Brassicaceae, suggest that the meristematic and vegetative function is shared between the *PNF* and *RPL* clades (Roeder et al., 2003; Smith et al., 2004; Konishi et al., 2006; Kanrar et al., 2008; Ung et al., 2011; Mühlhausen et al., 2013) and that it is likely the ancestral function of these genes in gymnosperm *RPL*-*PNF* homologs (Pabón-Mora et al., 2014) for which expression or functional studies have not been performed yet. Moreover, it is likely that all BLH proteins are involved in vegetative development, as it is a function that has been broadly described for other paralogs in Arabidopsis (Yu et al., 2009), such as *ARABIDOPSIS THALIANA* *HOMEBOX 1* (*ATH1*) that is expressed in the SAM and leaf primordia (Proveniers et al., 2007) and *SAWTOOTH1* (*SAW1* = *BLH2*) and *SAW2* (= *BLH5*) that regulate leaf development (Kumar et al., 2007).

***RPL* Homologs in Papaveraceae Show Broad Expression Patterns During Flower Development and More Restricted Expression During Fruit Development**

The floral organ expression patterns of *RPL* copies in the two Papaveraceae species (**Figures 5–7, 9**) are consistent with the expression patterns of *RPL* homologs in Brassicaceae (Roeder et al., 2003; Kanrar et al., 2006; Mühlhausen et al., 2013). Later in the developing carpel, *RPL* is mostly restricted to the stylar and stigmatic adaxial region of *B. frutescens* (**Figures 5–7**) similar to the expression described in Arabidopsis (Simonini et al., 2018). In general, the orthologs *PsomRPL* and *BofrRPL1*, members of the *RanRPL1* clade are similarly expressed in late stages of carpel development and in the fruit (**Figures 5, 9**), supporting also some degree of subfunctionalization between the *Bocconia* homologs that belong to the different clades. In Arabidopsis, *RPL* participates in the genetic network involved in the proper development of the style by interacting with auxin response factor *ETTIN* (*ETT*), *INDEHISCENT* (*IND*) and *BREVIPEDICELUS* (*BP*) (Marsch-Martinez et al., 2014; Simonini et al., 2016, 2018). Although *IND* is the result of a Brassicaceae specific duplication event that also gave rise to *HECATE3* genes, pre-duplication *HECATE3*-like genes are likely to maintain the same role in specifying the distal-most portion of the gynoecium (Pfannebecker et al., 2017; Gaillochet et al., 2018). The presence of *BP* orthologs in basal eudicots, known as *KNAT1* (Groot et al., 2005), together with our expression results point to a key conserved role of *RPL* in style development and proper carpel development, as the role of *RPL* is strictly to repress valve and valve margin developmental genes indirectly inducing replum formation (Alonso-Cantabrana et al., 2007) and in the maintenance of the meristematic activity during plant

development, specifically in the fruit (reviewed in Girin et al., 2009).

Of particular interest are the expression patterns of *RPL* in the fruit where dehiscence will occur, whether it is poricidal, in between the carpel central bundles as in *P. somniferum*, or opercular, in between the carpel margins and the commissural ring as in *B. frutescens*. This is suggestive of a role in specifying the separation layer and not the persistent tissue as in Arabidopsis. It is important to notice that even though we used a variety of *P. somniferum* where the fruits do not open, it did not interfere with the *RPL* expression in the putative separation layer of the fruit. The role in replum development may be Arabidopsis specific (Roeder et al., 2003), as it has not been found in other Brassicaceae (Mühlhausen et al., 2013). *B. frutescens* and *A. thaliana* present a similar fruit morphology, with a persistent medial tissue where the seeds remain attached after the two valves fall apart, and some of the genetic mechanisms involved in the fruit development are conserved such as the role of *SPT/ALC* genes in the specification of carpel margins and the dehiscence zone (Groszmann et al., 2011; Zumajo-Cardona et al., 2017) but the proper development of the persistent medial tissue is not determined by the same mechanisms in basal eudicots.

Functional analyses of *RPL* will help us to better understand their contribution to the diversification of fruits within Papaveraceae (**Figure 1**). Our expression data support the idea that although *RPL* is active during fruit development, its function in the maintenance of a persistent medial tissue is not conserved in basal eudicots. The replum in Arabidopsis seems to be the result of the co-option of *RPL*. During Arabidopsis carpel and fruit development, *RPL* is directly repressed by *APETALA2* (Ripoll et al., 2011) and *RPL*, negatively regulates *SHATERPROOF1/2* (*SHP1/2*), which are the paralogs of *AGAMOUS* (*AG*), to the replum boundary (Roeder et al., 2003; Kramer et al., 2004; Zahn et al., 2006; Chávez-Montes et al., 2015). Although no *SHP* orthologs have been found in basal eudicots (Pabón-Mora et al., 2014) it has been suggested that its function in fruit development is maintained by its ancestral gene *paleoAGAMOUS* (Hands et al., 2011). This regulatory network becomes more difficult to understand in *B. frutescens* due to its multiple copies. Nevertheless, the opposite expression patterns of the two copies of *AP2* in *Bocconia* (Zumajo-Cardona et al., 2017) with the expression of the three *RPL* copies presented here, suggest that the interaction of *AP2*-*RPL* is also maintained in this species. Thus, based on the protein analysis of *RPL* homologs and the expression patterns related to other genes in the network, it is likely that the interactions between these genes are conserved in Papaveraceae. In addition, we determined that *RPL* in Papaveraceae is expressed in the regions where the carpels distally fuse and subsequently in the separation layer of the fruits. Expression and functional analyses are required in other Ranunculales like Ranunculaceae (i.e., *Aquilegia coerulea*) or Menispermaceae (i.e., *Menispermum canadense*) in order to better understand the role of *RPL* in apocarpous gynoecium as well as the impact of the protein composition as we have detected that *RPL* sequences in these species are highly variable.

Finally, we described for the first time expression of *RPL* homologs in the developing ovules (**Figures 5–7, 9**). *BELL1*, also a TALE-Homeodomain gene closely related to *RPL* (Becker et al., 2002; Bowman et al., 2016) has been described to function in ovule development. *BELL1* represses *AGAMOUS* (*AG*) in the floral meristem (Bowman et al., 1991a,b; Bao et al., 2004). When *BELL1* is silenced, the continuous expression of *AG* in the ovule results in homeotic transformation of the integuments into carpels (Modrusan et al., 1994; Ray et al., 1994; Reiser et al., 1995; Hands et al., 2011). Later during ovule formation, *BELL1* is involved in the proper development of the integuments. The expression of *RPL* in the ovules of the two Papaveraceae species compared here (**Figures 5–7, 9**), particularly in *BofrRPL2* which seems to be specifically expressed in the early formation of the integuments, suggests that *RPL* genes may have the same role in ovule and proper integument development in basal eudicots (**Figure 6**). Here, we present for the first time also, the expression of *BofrRPL3* in the aril of *B. frutescens* (**Figure 7K**). Additional expression and functional analyses in non-model species that develop an aril are required in order to determine the conservation of this expression and the role of *RPL* during aril development.

AUTHOR CONTRIBUTIONS

All authors planned and designed the research, performed the experiments, analyzed the data, and wrote the final version of the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was funded in part by The Eppley Foundation for Research, Inc. (New York, NY, United States), by

COLCIENCIAS (Grant No. 111565842812) and by Convocatoria de Sostenibilidad- 2018–2019 and Convocatoria Programáticas 2017–2018 Universidad de Antioquia to the Grupo Evo-Devo en Plantas.

ACKNOWLEDGMENTS

We thank J. F. Alzate (Centro Nacional de Secuenciación de Genómica, SIU, Universidad de Antioquia, Medellín, Antioquia, Colombia) for the assembly and storage of our own generated transcriptomes.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.01833/full#supplementary-material>

FIGURE S1 | BofrRPL1, 2, 3 and PsomRPL protein sequences showing the regions where specific primers were designed. BofrRPL reverse primers were designed on the 3'UTR.

FIGURE S2 | MEME analysis showing conserved motifs across basal eudicots RPL protein sequences. Letter size denotes the degree of conservation of each amino acid.

TABLE S1 | List of the genes included in the phylogenetic analysis, with corresponding species, family and accession number of the sequences.

TABLE S2 | Primers used for the *in situ* hybridization analyses in the four RPL homologs.

TABLE S3 | Developmental landmarks for each stage of flower and fruit development in *Papaver somniferum*, based on Drea et al. (2007) and our complementary stages.

REFERENCES

- Alonso-Cantabrana, H., Ripoll, J. J., Ochando, I., Vera, A., Ferrándiz, C., and Martínez-Laborda, A. (2007). Common regulatory networks in leaf and fruit patterning revealed by mutations in the Arabidopsis ASYMMETRIC LEAVES1 gene. *Development* 134, 2663–2671. doi: 10.1242/dev.02864
- Alvarez, J., and Smyth, D. R. (2002). CRABS CLAW and SPATULA genes regulate growth and pattern formation during gynoecium development in *Arabidopsis thaliana*. *Int. J. Plant Sci.* 163, 17–41. doi: 10.1086/324178
- Ambrose, B. A., Lerner, D. R., Ciceri, P., Padilla, C. M., Yanofsky, M. F., and Schmid, R. J. (2000). Molecular and genetic analyses of the *silky1* gene reveal conservation in floral organ specification between eudicots and monocots. *Mol. Cell* 5, 569–579. doi: 10.1016/S1097-2765(00)80450-5
- Andrés, F., Romera Branchat, M., Martínez-Gallegos, R., Patel, V., Schneeberger, K., Jang, S., et al. (2015). Floral induction in Arabidopsis by FLOWERING LOCUS T requires direct repression of BLADE-ON-PETIOLE genes by the homeodomain protein PENNYWISE. *Plant Physiol.* 169, 2187–2199. doi: 10.1104/pp.15.00960
- Angiosperm Phylogeny Group (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* 181, 1–20. doi: 10.1111/boj.12385
- Arango-Ocampo, C., González, F., Alzate, J. F., and Pabón-Mora, N. (2016). The developmental and genetic bases of apetaly in *Bocconia frutescens* (Chelidoniaceae: Papaveraceae). *Evodevo* 7:16. doi: 10.1186/s1322701600546
- Argout, X., Salse, J., Aury, J., Guiltinan, M., Droc, G., Gouzy, J., et al. (2011). The genome of *Theobroma cacao*. *Nat. Genet.* 43, 101–108. doi: 10.1038/ng.736
- Arnaud, N., Lawrenson, T., Østergaard, L., and Sablowski, R. (2011). The same regulatory point mutation changed seed-dispersal structures in evolution and domestication. *Curr. Biol.* 21, 1215–1219. doi: 10.1016/j.cub.2011.06.008
- Arnaud, N., and Pautot, V. (2014). Ring the BELL and tie the KNOX: roles for TALEs in gynoecium development. *Front. Plant Sci.* 5:93. doi: 10.3389/fpls.2014.00093
- Avino, M., Kramer, E. M., Donohue, K., Hammel, A. J., and Hall, J. C. (2012). Understanding the basis of a novel fruit type in Brassicaceae: conservation and deviation in expression patterns of six genes. *Evodevo* 3:20. doi: 10.1186/2041-9139-3-20
- Bailey, T. L., Williams, N., Mischak, C., and Li, W. W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* 34, 369–373. doi: 10.1093/nar/gkl198
- Bao, X., Franks, R. G., Levin, J. Z., and Liu, Z. (2004). Repression of *AGAMOUS* by *BELLRINGER* in floral and inflorescence meristems. *Plant Cell* 16, 1478–1489. doi: 10.1105/tpc.021147
- Becker, A., Bey, M., Burglin, T. R., Saedler, H., and Theissen, G. (2002). Ancestry and diversity of BEL1-like homeobox genes revealed by gymnosperm (*Gnetum gnemon*) homologs. *Dev. Genes Evol.* 212, 452–457. doi: 10.1007/s00427-002-0259-7
- Becker, A., Gleissberg, S., and Smyth, D. R. (2005). Floral and vegetative morphogenesis in *California poppy* (*Eschscholzia californica* CHAM.). *Int. J. Plant Sci.* 166, 537–555. doi: 10.1086/429866

- Bencivenga, S., Serrano-Mislata, A., Bush, M., Fox, S., and Sablowski, R. (2016). Control of oriented tissue growth through repression of organ boundary genes promotes stem morphogenesis. *Dev. Cell* 39, 198–208. doi: 10.1016/j.devcel.2016.08.013
- Bhatt, A. M., Etchells, J. P., Canales, C., Lagodienko, A., and Dickinson, H. (2004). VAAMANA-a BEL1-like homeodomain protein, interacts with KNOX proteins BP and STM and regulates inflorescence stem growth in Arabidopsis. *Gene* 328, 103–111. doi: 10.1016/j.gene.2003.12.033
- Bowman, J. L., Drews, G. N., and Meyerowitz, E. M. (1991a). Expression of the Arabidopsis floral homeotic gene AGAMOUS is restricted to specific cell types late in flower development. *Plant Cell* 3, 749–758.
- Bowman, J. L., Smyth, D. R., and Meyerowitz, E. M. (1991b). Genetic interactions among floral homeotic genes of Arabidopsis. *Development* 112, 1–20.
- Bowman, J. M., Sakakibara, K., Furumizu, C., and Dierschke T. (2016). The evolution of the cycle of life. *Annu. Rev. Genet.* 50, 133–154. doi: 10.1146/annurev-genet-120215-035227
- Bürglin, T. R. (1997). Analysis of TALE super-class homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals. *Nucleic Acids Res.* 25, 4173–4180. doi: 10.1093/nar/25.21.4173
- Byrne, M., Groover, A. T., Fontana, J. R., and Martienssen, R. A. (2003). Phyllotactic pattern and stem cell fate are determined by the Arabidopsis homeobox gene BELLRINGER. *Development* 130, 3941–3950. doi: 10.1242/dev.00620
- Campbell, B. A., Hallengren, J., and Hannapel, D. J. (2008). Accumulation of BEL1-like transcripts in *Solanaceous* species. *Planta* 228, 897–906. doi: 10.1007/s00425-008-0780-7
- Chan, R. L., Gago, G. M., Palena, C. M., and Gonzalez, D. H. (1998). Homeo-boxes in plant development. *Biochim. Biophys. Acta* 1442, 1–19. doi: 10.1016/S0167-4781(98)00119-5
- Chávez-Montes, R. A., Herrera-Ubaldo, H., Serwatowska, J., and de Folter, S. (2015). Towards a comprehensive and dynamic gynoecium gene regulatory network. *Curr. Plant Bio.* 3, 3–12. doi: 10.1016/j.cpb.2015.08.002
- Chung, K. S., Lee, J. H., Lee, J. S., and Ahn, J. H. (2013). Fruit indehiscence caused by enhanced expression of NO TRANSMITTING TRACT in *Arabidopsis thaliana*. *Mol. Cell* 35, 519–525. doi: 10.1007/s10059-013-0030-0
- Cole, M., Nolte, C., and Werr, W. (2006). Nuclear import of the transcription factor SHOOT MERISTEMLESS depends on heterodimerization with BLH proteins expressed in discrete sub-domains of the shoot apical meristem of *Arabidopsis thaliana*. *Nucleic Acids Res.* 34, 1281–1292. doi: 10.1093/nar/gkl016
- Davie, J. H. (1935). Chromosome studies in the malvaceae and certain related families. *Genetica* 17, 487–498. doi: 10.1007/BF01508190
- D'Hont, A., Denoeud, F., Aury, J. M., Baurens, F. C., Carreel, F., Garsmeur, O., et al. (2012). The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* 488, 213–217. doi: 10.1038/nature11241
- Dinneny, J. R., Weigel, D., and Yanofsky, M. F. (2005). A genetic framework for fruit patterning in *Arabidopsis thaliana*. *Development* 132, 4687–4696. doi: 10.1242/dev.02062
- Drea, S., Hileman, L. C., de Martino, G., and Irish, V. F. (2007). Functional analyses of genetic pathways controlling petal specification in poppy. *Development* 134, 4157–4166. doi: 10.1242/dev.013136
- Ferrandiz, C., Pelaz, S., and Yanofsky, M. F. (1999). Control of carpel and fruit development in Arabidopsis. *Annu. Rev. Biochem.* 99, 321–354. doi: 10.1146/annurev.biochem.68.1.321
- Fourquin, C., and Ferrandiz, C. (2012). Functional analyses of AGAMOUS family members in *Nicotiana benthamiana* clarify the evolution of early and late roles of C-function genes in eudicots. *Plant J.* 71, 990–1001. doi: 10.1111/j.1365-313X.2012.05046.x
- Gaillochet, C., Jamge, S., van der Wal, F., Angenent, G., Immink, R., and Lohmann, J. U. (2018). A molecular network for functional versatility of HECATE transcription factors. *Plant J.* 95, 57–70. doi: 10.1111/tpj.13930
- Gasser, C. S., and Simon, M. K. (2011). Seed dispersal: same gene, different organs. *Curr. Biol.* 21, 546–548. doi: 10.1016/j.cub.2011.06.035
- Girin, T., Sorefan, K., and Ostergaard, L. (2009). Meristematic sculpting in fruit development. *J. Exp. Bot.* 60, 1493–1502. doi: 10.1093/jxb/erp031
- Groot, E. P., Sinha, N., and Gleissberg, S. (2005). Expression patterns of STM-like KNOX and histone H4 genes in shoot development of the dissected-leaved basal eudicot plants *Chelidonium majus* and *Eschscholzia californica* (Papaveraceae). *Plant Mol. Biol.* 58, 317–331. doi: 10.1007/s11103-005-4548-1
- Groszmann, M., Paicu, T., Alvarez, J. P., Swain, S. M., and Smyth, D. R. (2011). SPATULA and ALCATRAZ, are partially redundant, functionally diverging bHLH genes required for Arabidopsis gynoecium and fruit development. *Plant J.* 68, 816–829. doi: 10.1111/j.1365-313X.2011.04732.x
- Gunn, C. R. (1980). Seeds and fruits of papaveraceae and fumariaceae. *Seed Sci. Technol.* 8, 4–58.
- Hake, S., Smith, H. M. S., Holtan, H., Magnani, E., Mele, G., and Ramirez, J. (2004). The role of KNOX genes in plant development. *Annu. Rev. Cell Dev. Biol.* 20, 125–151. doi: 10.1146/annurev.cellbio.20.031803.093824
- Hands, P., Vosnakis, N., Betts, D., Irish, V. F., and Drea, S. (2011). Alternate transcripts of a floral developmental regulator have both distinct and redundant functions in opium poppy. *Ann. Bot.* 107, 1557–1566. doi: 10.1093/aob/mcr045
- Hay, A., and Tsiantis, M. A. (2009). KNOX family TALE. *Curr. Opin. Plant Biol.* 12, 593–598. doi: 10.1016/j.pbi.2009.06.006
- Jain, R., and Prasad, B. (2014). Karyotype analysis and chromosome number confirmation in *Tinospora cordifolia* (willd.) miers. ex Hook. f. and Thomas. *A. Int. J. Pharm. Bio Sci.* 5, 996–1001.
- Jiao, Y., Wickett, N. J., Ayyampalayam, S., Chanderbali, A. S., Landherr, L., and Ralph, P. E., et al. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature* 473, 97–100. doi: 10.1038/nature09916
- Kaderet, J. W. (1993). “Papaveraceae,” in *The Families and Genera of Vascular Plants*, ed. K. Kubitzki (Berlin: Springer), 494–506.
- Kanrar, S., Bhattacharya, M., Arthur, B., Courtier, J., and Smith, H. M. (2008). Regulatory networks that function to specify flower meristems require the function of homeobox genes PENNYWISE and POUND-FOOLISH in Arabidopsis. *Plant J.* 54, 924–937. doi: 10.1111/j.1365-313X.2008.03458.x
- Kanrar, S., Onguka, O., and Smith, H. M. S. (2006). Arabidopsis inflorescence architecture requires the activities of KNOX-BELL homeodomain heterodimers. *Planta* 224, 1163–1173. doi: 10.1007/s00425-006-0298-9
- Kapoor, L. D. (1995). *Opium Poppy: Botany, Chemistry, and Pharmacology*. Binghamton, NY: Haworth Press.
- 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/gk436
- Khan, M., Ragni, L., Tabb, P., Salasini, B. C., Chatfield, S., Datla, R., et al. (2015). Repression of lateral organ boundary genes by PENNYWISE and POUND-FOOLISH is essential for meristem maintenance and flowering in Arabidopsis. *Plant Physiol.* 169, 2166–2186. doi: 10.1104/pp.15.00915
- Khan, M., Tabb, P., and Hepworth, S. R. (2012). BLADE-ON-PETIOLE1 and 2 regulate Arabidopsis inflorescence architecture in conjunction with homeobox genes KNAT6 and ATH1. *Plant Signal. Behav.* 7, 788–792. doi: 10.4161/psb.20599
- Kim, S., Soltis, D. E., Soltis, P. S., Zanis, M. J., and Suh, Y. (2004). Phylogenetic relationships among early-diverging eudicots based on four genes: were the eudicots ancestrally woody? *Mol. Phylogenet. Evol.* 31, 16–30. doi: 10.1016/j.ympev.2003.07.017
- Konishi, S., Izawa, T., Lin, S., Ebana, K., Fukuta, Y., Sasaki, T., et al. (2006). An SNP caused loss of seed shattering during rice domestication. *Nature* 312, 1392–1396. doi: 10.1126/science.1126410
- Kramer, E. M., Jaramillo, M. A., and Di Stilio, V. S. (2004). Patterns of gene duplication and functional evolution during the diversification of the AGAMOUS subfamily of MADS box genes in angiosperms. *Genetics* 166, 1011–1023. doi: 10.1534/genetics.166.2.1011
- Kraus, J. E., De Souza, H. C., Rezende, M. H., Castro, N. M., Vecchi, C., and Luque, R. (1998). Astra blue and basic fuchsin double staining of plant material. *Biotech. Histochem.* 73, 235–243. doi: 10.3109/10520299809141117
- Kumar, R., Kushalappa, K., Godt, D., Pidkowich, M. S., Pastorelli, S., Hepworth, S. R., et al. (2007). The Arabidopsis BEL1-LIKE HOMEODOMAIN proteins SAW1 and SAW2 act redundantly to regulate KNOX expression spatially in leaf margins. *Plant Cell* 19, 2719–2735. doi: 10.1105/tpc.106.048769
- Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large data sets. *Bioinformatics* 22, 3276–3278. doi: 10.1093/bioinformatics/btu531
- Lin, Z., Griyith, M. E., Li, X., Zhu, Z., Tan, L., Fu, Y., et al. (2007). Origin of seed shattering in rice (*Oryza sativa* L.). *Planta* 226, 11–20. doi: 10.1007/s00425-006-0460-4
- March-Martinez, N., Zuniga-Mayo, V. M., Herrera-Ubaldo, H., Ouwerkerk, P. B., Pablo-Villa, J., Lozano-Sotomayor, P., et al. (2014). The NTT transcription

- factor promotes replum development in Arabidopsis fruits. *Plant J.* 80, 69–81. doi: 10.1111/tpj.12617
- Meyer, R. S., and Purugganan, M. D. (2013). Evolution of crop species: genetics of domestication and diversification. *Nat. Rev. Genet.* 14, 840–852. doi: 10.1038/nrg3605
- Miller, M. A., Pfeiffer, W., and Schwartz, T. (2010). “Creating the CIPRES Science Gateway for inference of large phylogenetic trees,” in *Proceedings of the Gateway Computing Environments Workshop (GCE)*, New Orleans, LA, 1–8.
- Modrusan, Z., Reiser, L., Feldmann, K. A., Fischer, R. L., and Haughn, G. W. (1994). Homeotic transformation of ovules into carpel-like structures in Arabidopsis. *Plant Cell* 6, 333–349. doi: 10.1105/tpc.6.3.333
- Mühlhausen, A., Lenser, T., Mummenhoff, K., and Theissen, G. (2013). Evidence that an evolutionary transition from dehiscent to indehiscent fruits in *Lepidium* (Brassicaceae) was caused by a change in the control of valve margin identity genes. *Plant J.* 73, 824–835. doi: 10.1111/tpj.12079
- Mukherjee, K., Brocchieri, L., and Bürglin, T. R. (2009). A comprehensive classification and evolutionary analysis of plant homeobox genes. *Mol. Biol. Evol.* 26, 2775–2794. doi: 10.1093/molbev/msp201
- Nicholls, A., Sharp, K. A., and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* 11, 281–296. doi: 10.1002/prot.340110407
- Ortiz-Ramírez, C., Plata-Arboleda, S., and Pabón-Mora, N. (2018). Evolution of genes associated with gynoeceum patterning and fruit development in Solanaceae. *Ann. Bot.* 121, 1211–1230. doi: 10.1093/aob/mcy007
- Pabón-Mora, N., Wong, K. G., and Ambrose, B. (2014). Evolution of fruit development genes in flowering plants. *Front. Plant Sci.* 5:300. doi: 10.3389/fpls.2014.00300
- Panchy, N., Lehti-Shiu, M., and Shiu, S. H. (2016). Evolution of gene duplication in plants. *Plant Physiol.* 171, 2294–2316. doi: 10.1104/pp.16.00523
- Pfannebecker, K. C., Lange, M., Rupp, O., and Becker, A. (2017). Seed plant specific gene lineages involved in carpel development. *Mol. Biol. Evol.* 34, 925–942. doi: 10.1093/molbev/msw297
- Proveniers, M., Rutjens, B., Brand, M., and Smeekens, S. (2007). The Arabidopsis TALE homeobox gene ATH1 controls floral competency through positive regulation of FLC. *Plant J.* 52, 899–913. doi: 10.1111/j.1365-313X.2007.03285.x
- Ray, A., Robinson-Beers, K., Ray, S., Baker, S. C., Lang, J. D., Preuss, D., et al. (1994). Arabidopsis floral homeotic gene BELL (BEL1) controls ovule development through negative regulation of AGAMOUS gene (AG). *Proc. Natl. Acad. Sci. U.S.A.* 91, 5761–5765. doi: 10.1073/pnas.91.13.5761
- Reiser, L., Modrusan, Z., Margossian, L., Samach, A., Ohad, N., Haughn, G. W., et al. (1995). The *BELL1* gene encodes a homeodomain protein involved in pattern formation in the Arabidopsis ovule primordium. *Cell* 83, 735–742. doi: 10.1016/0092-8674(95)90186-8
- Reyes-Olalde, J. I., Zuniga-Mayo, V. M., Chavez Montes, R. A., Marsch-Martinez, N., and de Folter, S. (2013). Inside the gynoeceum: at the carpel margin. *Trends Plant Sci.* 11, 644–655. doi: 10.1016/j.tplants.2013.08.002
- Ripoll, J. J., Roeder, A. H. K., Ditta, G. S., and Yanofsky, M. F. (2011). A novel role for the floral homeotic gene APETALA2 during Arabidopsis fruit development. *Development* 138, 5167–5176. doi: 10.1242/dev.073031
- Roeder, A. H. K., Ferrándiz, C., and Yanofsky, M. F. (2003). The role of the *REPLUMLESS* homeodomain protein in patterning the Arabidopsis Fruit. *Curr. Biol.* 13, 1630–1635. doi: 10.1016/j.cub.2003.08.027
- Roeder, A. H. K., and Yanofsky, M. F. (2006). Fruit development in Arabidopsis. *Arabidopsis Book* 4:e0075. doi: 10.1199/tab.0075
- Roth, I. (1977). *Fruits of Angiosperms*. Berlin: Gebr. Borntraeger.
- Rutjens, B., Bao, D., van Eck-Stouten, E., Brand, M., Smeekens, S., and Proveniers, M. (2009). Shoot apical meristem function in Arabidopsis requires the combined activities of three BEL1-like homeodomain proteins. *Plant J.* 58, 641–654. doi: 10.1111/j.1365-313X.2009.03809.x
- Sárkány, S., and Szalay, I. (1964). *Növénytani Praktikum I. Növényismeret*. Budapest: Tankönyvkiadó.
- Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., et al. (2010). Genome sequence of the palaeopolyploid soybean. *Nature* 463, 178–183. doi: 10.1038/nature08670
- Simonini, S., Deb, J., Moubayidin, L., Stephenson, P., Valluru, M., Freire-rios, A., et al. (2016). A noncanonical auxin-sensing mechanism is required for organ morphogenesis in Arabidopsis. *Genes Dev.* 30, 2286–2296. doi: 10.1101/gad.285361.116
- Simonini, S., Stephenson, P., and Østergaard, L. (2018). A molecular framework controlling style morphology in Brassicaceae. *Development* 145:dev158105. doi: 10.1242/dev.158105
- Smith, H. M. S., Campbell, B. C., and Hake, S. (2004). Competence to respond to floral inductive signals requires the homeobox genes PENNYWISE and POUND-FOOLISH. *Curr. Biol.* 14, 812–817. doi: 10.1016/j.cub.2004.04.032
- Smith, H. M. S., and Hake, S. (2003). The interaction of two homeobox genes, BREVPEDICELLUS and PENNYWISE, regulates internode patterning in the Arabidopsis inflorescence. *Plant Cell* 15, 1717–1727. doi: 10.1105/tpc.012856
- Stamatakis, A., Hoover, P., and Rougemont, J. (2008). A rapid bootstrap algorithm for the RAxML Web servers. *Syst. Biol.* 57, 758–771. doi: 10.1080/10635150802429642
- Sugiura, T. (1940). Chromosome studies on Papaveraceae with special reference to the phylogeny. *Cytologia* 10, 558–576. doi: 10.1508/cytologia.10.558
- Ung, N., Lal, S., and Smith, H. M. (2011). The role of PENNYWISE and POUND-FOOLISH in the maintenance of the shoot apical meristem in Arabidopsis. *Plant Physiol.* 156, 605–614. doi: 10.1104/pp.110.171462
- Yu, L., Patibanda, V., and Smith, H. M. (2009). A novel role of BELL1-like homeobox genes, PENNYWISE and POUND-FOOLISH, in floral patterning. *Planta* 229, 693–707. doi: 10.1007/s00425-008-0867-1
- Zahn, L. M., Leebens-Mack, J. H., Arrington, J. M., Hu, Y., Landherr, L. L., dePamphilis, C. W., et al. (2006). Conservation and divergence in the AGAMOUS subfamily of MADS-box genes: evidence of independent sub- and neofunctionalization events. *Evol. Dev.* 8, 30–45. doi: 10.1111/j.1525-142X.2006.05073.x
- Zumajo-Cardona, C., Ambrose, B., and Pabon-Mora, N. (2017). Evolution of the SPATULA/ALCATRAZ gene lineage and expression analyses in the basal eudicot, *Bocconia frutescens* L. (Papaveraceae). *Evodevo* 8:5. doi: 10.1186/s13227-017-0068-8

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 © 2018 Zumajo-Cardona, Pabón-Mora and Ambrose. 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 and Diversification of *FRUITFULL* Genes in Solanaceae

Dinusha C. Maheepala¹, Christopher A. Emerling^{2†}, Alex Rajewski¹, Jenna Macon¹, Maya Strahl^{3†}, Natalia Pabón-Mora⁴ and Amy Litt^{1*}

OPEN ACCESS

Edited by:

Jill Christine Preston,
University of Vermont, United States

Reviewed by:

Renata Reinheimer,
Instituto de Agrobiotecnología del
Litoral (IAL), Argentina
Bharti Sharma,
California State Polytechnic University,
Pomona, United States

*Correspondence:

Amy Litt
amy.litt@ucr.edu

†Present address:

Christopher A. Emerling,
Whittier College, Whittier, CA,
United States
Maya Strahl,
Genetics and Genomic Sciences,
Icahn School of Medicine at Mount
Sinai, New York, NY, United States

Specialty section:

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

Received: 29 September 2018

Accepted: 11 January 2019

Published: 21 February 2019

Citation:

Maheepala DC, Emerling CA,
Rajewski A, Macon J, Strahl M,
Pabón-Mora N and Litt A (2019)
Evolution and Diversification
of *FRUITFULL* Genes in Solanaceae.
Front. Plant Sci. 10:43.
doi: 10.3389/fpls.2019.00043

¹ Department of Botany and Plant Sciences, University of California, Riverside, Riverside, CA, United States, ² Institut des Sciences de l'Évolution de Montpellier, Université de Montpellier, Centre National de la Recherche Scientifique, Institut de Recherche pour le Développement, École Pratique des Hautes Études, Montpellier, France, ³ The New York Botanical Garden, Bronx, NY, United States, ⁴ Instituto de Biología, Universidad de Antioquia, Medellín, Colombia

Ecologically and economically important fleshy edible fruits have evolved from dry fruit numerous times during angiosperm diversification. However, the molecular mechanisms that underlie these shifts are unknown. In the Solanaceae there has been a major shift to fleshy fruits in the subfamily Solanoideae. Evidence suggests that an ortholog of *FRUITFULL* (*FUL*), a transcription factor that regulates cell proliferation and limits the dehiscence zone in the silique of *Arabidopsis*, plays a similar role in dry-fruited Solanaceae. However, studies have shown that *FUL* orthologs have taken on new functions in fleshy fruit development, including regulating elements of tomato ripening such as pigment accumulation. *FUL* belongs to the core eudicot *euFUL* clade of the angiosperm *AP1/FUL* gene lineage. The *euFUL* genes fall into two paralogous clades, *euFULI* and *euFULII*. While most core eudicots have one gene in each clade, Solanaceae have two: *FUL1* and *FUL2* in the former, and *MBP10* and *MBP20* in the latter. We characterized the evolution of the *euFUL* genes to identify changes that might be correlated with the origin of fleshy fruit in Solanaceae. Our analyses revealed that the Solanaceae *FUL1* and *FUL2* clades probably originated through an early whole genome multiplication event. By contrast, the data suggest that the *MBP10* and *MBP20* clades are the result of a later tandem duplication event. *MBP10* is expressed at weak to moderate levels, and its atypical short first intron lacks putative transcription factor binding sites, indicating possible pseudogenization. Consistent with this, our analyses show that *MBP10* is evolving at a faster rate compared to *MBP20*. Our analyses found that Solanaceae *euFUL* gene duplications, evolutionary rates, and changes in protein residues and expression patterns are not correlated with the shift in fruit type. This suggests deeper analyses are needed to identify the mechanism underlying the change in *FUL* ortholog function.

Keywords: dry fruit, fleshy fruit, fruit development, fruit evolution, *FRUITFULL*, gene duplication, MADS-box transcription factors, Solanaceae

INTRODUCTION

Fleshy fruits are agriculturally and economically important plant organs that have evolved from dry fruits many times during angiosperm evolution. However, the genetic changes that are required for this shift to occur are as yet unknown (Bolmgren and Eriksson, 2010). In the agriculturally, pharmacologically, and horticulturally important plant family Solanaceae (nightshades), there was a shift to fleshy fruit in the subfamily Solanoideae from plesiomorphic dry fruit (**Figure 1**) (Knapp, 2002). In the family two independent transitions to fleshy fruits have also occurred in the genera *Duboisia* (subfamily Anthocercideae) and *Cestrum* (subfamily Cestroideae), as well as a reversal to dry fruit in the genus *Datura* (subfamily Solanoideae) (Knapp, 2002).

Evidence from tomato (*Solanum lycopersicum*, subfamily Solanoideae) indicates that *FRUITFULL* (*FUL*) transcription factors (TFs) have novel functions in fleshy fruit development compared to *Arabidopsis* (Brassicaceae) and *Nicotiana* (Solanaceae, subfamily Nicotianoideae) (Gu et al., 1998; Smykal et al., 2007; Bemer et al., 2012; Shima et al., 2013, 2014; Wang et al., 2014). *FUL* is a MADS-box TF that plays pleiotropic roles in both reproductive and vegetative development in the model plant *Arabidopsis thaliana* (Spence et al., 1996; Gu et al., 1998; Liljegren et al., 2000; Rajani and Sundaresan, 2001; Melzer et al., 2008). *FUL* controls cell proliferation in the fruit valves and spatially limits the formation of the dehiscence zone in the dry silique of *A. thaliana*, enabling the mature fruits to dehisce (Spence et al., 1996; Gu et al., 1998; Liljegren et al., 2000, 2004; Rajani and Sundaresan, 2001). Overexpression of a *Nicotiana tabacum* *FUL* ortholog in woodland tobacco (*Nicotiana sylvestris*) resulted in indehiscent fruits with reduced lignification at the dehiscence zones, suggesting a role similar to that observed in silique development in *A. thaliana* (Smykal et al., 2007). Several groups have examined the function of *euFUL* genes, the core-eudicot clade to which *FUL* belongs, in tomato (Bemer et al., 2012; Shima et al., 2014; Wang et al., 2014). All studies showed defects in fruit pigmentation during ripening when *FUL* ortholog expression was downregulated, and some studies also suggested roles in ethylene production and pericarp and cuticle thickness (Bemer et al., 2012; Shima et al., 2014; Wang et al., 2014). These data indicate that *euFUL* genes are controlling different processes in dry and fleshy fruits in the Solanaceae.

Early in the diversification of core-eudicots, there was a duplication in the *euFUL* gene clade, which resulted in the *euFULI* and *euFULII* clades (Litt and Irish, 2003; Shan et al., 2007). The *A. thaliana* *FUL* gene belongs to the *euFULI* clade while its paralog, *AGL79* which plays a role in lateral root development, branching, leaf morphology, and transition to flowering, belongs to the *euFULII* clade (Gao et al., 2018). The *euFULI* clade has duplicated in Solanaceae resulting in two subclades, designated here as *FUL1* and *FUL2*; likewise the *euFULII* clade has two Solanaceae-specific subclades, here designated *MBP10* and *MBP20* (Hileman et al., 2006; Bemer et al., 2012; The Tomato Genome Consortium, 2012). We studied the evolution of *euFUL* genes in Solanaceae to characterize patterns of selection, duplication, and sequence evolution to identify changes that might be correlated with the shift to fleshy fruit. We

tested the following hypotheses: (1) following the duplication of *euFUL* genes, there was a relaxation of selection in some or all of the resulting clades that resulted in sequence diversification; (2) changes in amino acid sequences are correlated with the origin of fleshy fruit. Although we found several sites showing changes in amino acid residues that might have resulted in changes in protein function, none of these were associated with the evolution of fleshy fruit. Consistent with our hypothesis, we found that the *FUL1* and *MBP10* genes are evolving at significantly faster rates in comparison to *FUL2* and *MBP20*. In combination with the relatively weak expression of *MBP10* and loss of potential regulatory elements, our data suggest that the *MBP10* lineage may be undergoing pseudogenization.

MATERIALS AND METHODS

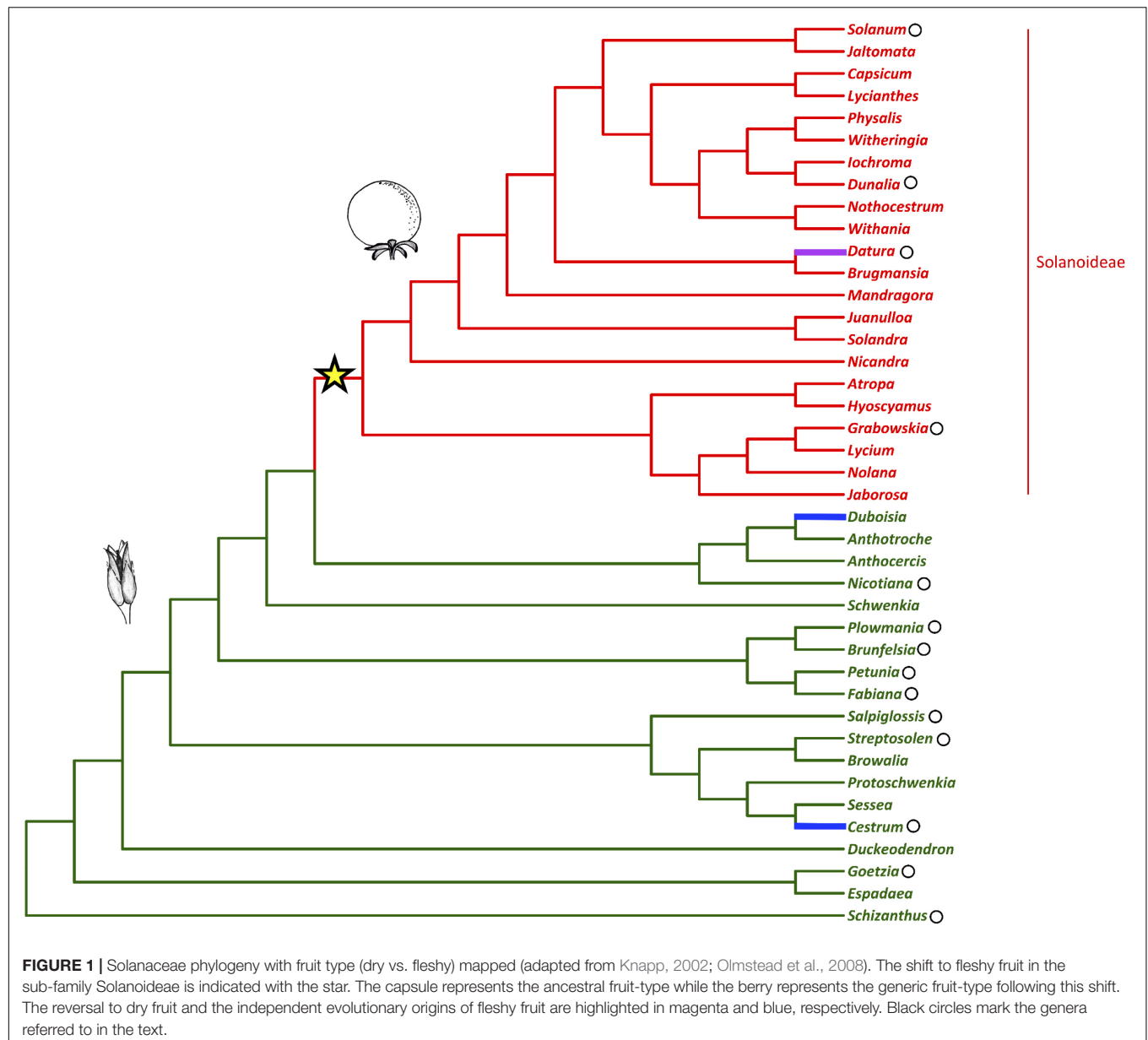
Plant Material for Sequencing

Sources of plant and tissue material for sequencing are listed in **Table S1** in **Supplementary Data Sheet S1**. Plants were grown in temperature controlled glasshouses at University of California, Riverside (UCR), The New York Botanical Garden, NY (NYBG), and The University of Antioquia, Colombia (UdeA) or collected from the grounds at UCR and the Universidad de Antioquia or the field at Parque Arvi, Vereda Santa Elena, El Tambo, Colombia.

For ease of reference and to simplify language, throughout the paper, members of Solanoideae, including the dry-fruited *Datura*, will be referred to as “fleshy-fruited species” (rather than “fleshy-fruited species and *Datura*”). Likewise non-Solanoideae, including the fleshy-fruited *Cestrum* and *Duboisia*, will be referred to as “dry-fruited species” (rather than “dry-fruited species and *Cestrum* and *Duboisia*”).

RNA Isolation, cDNA Synthesis/Library Preparation, and Sequencing

RNA was extracted from fruit, floral/inflorescence or leaf tissue using RNeasy Plant Mini Kits (QIAGEN, Hilden, Germany) according to the manufacturer's protocol. For *Grabowskia glauca*, *Dunalia spinosa*, *Fabiana viscosa*, and *Salpiglossis sinuata* RNA extractions, lysis buffer RLC was used instead of RLT and 2.5% (w/v) polyvinylpyrrolidone (PVP) was added. The RLT buffer was used for extracting RNA from all other species. RNA quality was checked using a BioSpectrometer Basic (Eppendorf, Hamburg, Germany) and stored at -80°C . cDNA was synthesized using SuperScript III Reverse Transcriptase (Thermo Fisher, San Diego, CA, United States) according to the manufacturer's protocol and the product was checked by amplifying *ACTIN*. Clade-specific degenerate primers were designed to target specific *euFUL* gene homologs based on conserved regions in Solanaceae *euFUL* gene alignments (**Supplementary Table S5**). PCR was run for two initial cycles with an annealing temperature between 40 and 45°C followed by 30 cycles at 55°C annealing temperature. The PCR products were visualized on a 1% agarose gel. If multiple amplicon band sizes were present, the annealing temperature of the first two cycles was increased until only one product size was achieved.



PCR products were purified using QIAquick PCR Purification Kit (QIAGEN) according to the manufacturer's protocol. The purified product was then cloned using TOPO TA Cloning Kit (Life Technologies, Carlsbad, CA, United States) according to the manufacturer's protocol, and the ligated plasmids were transformed into chemically competent TOP10 strain of *Escherichia coli*. Transformants were plated on LB plates with kanamycin selection (50 µg/mL) coated with 40 µL of 25 mg/mL X-Gal and IPTG, and incubated at 37°C overnight. Individual positive (white) colonies were used as templates in amplification with M13F and M13R primers (Life Technologies, Carlsbad, CA, United States) to identify those colonies with inserts of the expected size between 500 bp and 1 kb. These were grown overnight in 5 mL liquid LB medium supplemented with kanamycin (50 µg/mL) in an incubator-shaker at 250

RPM and 37°C. Plasmids were extracted from the liquid cultures using Plasmid Miniprep Kit (QIAGEN) according to the manufacturer's protocol, and sequenced using M13 reverse primer at the Institute for Integrative Genome Biology (IIGB) at UCR or Eton Bioscience, Inc. (San Diego, CA, United States).

For library preparation, RNA quality was checked using a Bioanalyzer (Agilent, Santa Clara, CA, United States). RNAseq library preparation was done according to the manufacturer's Poly(A) mRNA Magnetic Isolation Module protocol for NEBNext Ultra Directional RNA library Prep Kit for Illumina (New England BioLabs, Ipswich, MA, United States). *Cestrum diurnum*, *C. nocturnum*, and *Schizanthus grahamii* libraries were sequenced on an Illumina NextSeq v2 platform with high-output runs of 75 bp paired-end reads while *Dunalia spinosa*, *Fabiana viscosa*, *Grabowskia glauca*, and *Salpiglossis*

sinuata libraries were sequenced on an Illumina NextSeq v2 platform with high-output runs of both 75 bp paired-end reads and 150 bp single-end reads at IIGB, UCR. *Nicotiana obtusifolia* libraries were generated at NYBG and sequenced at the Beijing Genomics Institute (Shenzhen, China), and *Brunfelsia australis* and *Streptosolen jamesonii* libraries (Ortiz-Ramírez et al., 2018) were generated at UdeA and sequenced at Macrogen (Korea). All resulting *euFUL* sequences from both degenerate primer PCRs and transcriptomes are listed in **Table S2** in **Supplementary Data Sheet S1**. Individual sequences from PCR-based methods have been deposited in the GenBank (for accession numbers, see **Table S1** in **Supplementary Data Sheet S1**) and transcriptome data for *N. obtusifolia*, *C. diurnum*, *C. nocturnum*, *D. spinosa*, *F. viscosa*, *G. glauca*, and *S. grahamii* have been deposited on the SolGenomics network¹.

Mining *euFUL* Sequences From *de novo* Transcriptome Assembly and Databases

For transcriptome assembly, raw paired-end reads and single-end reads from Illumina sequencing were first quality trimmed using Trimmomatic v0.36 (Bolger et al., 2014) or TrimGalore (Krueger, 2017) and *de novo* assembled on the UCR High Performance Computing Cluster (HPCC) using the default settings of Trinity v2.4.0 (Grabherr et al., 2011). *Dunalia spinosa*, *Fabiana viscosa*, *Grabowskia glauca*, and *Salpiglossis sinuata* libraries were assembled by combining both 75 bp paired-end and 150 bp single-end reads. Each assembled transcriptome was then used to create a custom Basic Local Alignment Tool (BLAST) (Altschul et al., 1990) database. The BLAST database for each species was queried on the HPCC with both blastn and tblastx using all available sequences in our *euFUL* sequence file using a UNIX command line that sequentially matched each sequence in our query file against the database (BLAST® Command Line Applications User Manual, 2008). BLAST analyses were also conducted on the NCBI² (NCBI Resource Coordinators, 2017) and oneKP³ (Matasci et al., 2014) databases using *A. thaliana* *FUL* and various Solanaceae *FUL* homologs as query. Matching output sequences (**Table S1** in **Supplementary Data Sheet S1**) from both transcriptomes assemblies and database mining were further confirmed by compiling a gene tree as described below. We confirmed the accuracy of our sequences using gene specific primers and Sanger sequencing. Unless specified otherwise, all sequences referred to in this manuscript are the full or partial mRNA sequences.

Gene-Tree Generation

The Multiple Sequence Comparison by Log-Expectation (MUSCLE) (Edgar, 2004) tool was used to align *euFUL* sequences (**Supplementary Data Sheet S2**). The appropriate model for tree building, GTR+G, was determined with jModelTest 2.0 (Darriba et al., 2012). Ten independent maximum likelihood (ML) analyses starting with random trees were performed using GARLI v2.1 (Genetic Algorithm for Rapid Likelihood Inference)

(Bazin et al., 2014). *euFUL* genes from Convolvulaceae (*Convolvulus*, *Cuscuta* and *Ipomoea* species), which were retrieved from the oneKP database⁴, were designated as the outgroup in each analysis, which meant these sequences were automatically excluded from the ingroup clades. Each ML run was set to terminate when there was no significantly better scoring topology for 20,000 consecutive generations. The ten resulting trees were checked for agreement by calculating the pairwise Robinson–Foulds distance using ‘ape’ and ‘phangorn’ packages on R (Robinson and Foulds, 1981; Paradis et al., 2004; Schliep, 2010; R Core Team, 2018). The tree with the largest ML value was chosen as the starting tree in a bootstrap analysis involving 1,000 replicates. The results of the replicates were summarized and bootstrap values were calculated using SumTrees tool of DendroPy package on Python ver. 2.7 (Python Language Reference, 2010; Sukumaran and Holder, 2010) or Geneious 10.2 (Darling et al., 2010; Kearsse et al., 2012).

Any sequences that did not group with any of the subclades were aligned with the paralogs to investigate whether these may have been splice isoforms. Any such isoform was expected to have large insertions/deletions at splice junctions. None were noted.

Selection Pressure Analysis

The CODEML program within the Phylogenetic Analysis by Maximum Likelihood (PAML) (Yang, 1997) v 1.3⁵ software package was run on the HPCC at UCR to analyze the selection pressure acting on *euFUL* genes. These analyses were performed to test if different gene lineages as well as sub-groups within those lineages were evolving at significantly different rates. Further scenarios were considered in which each gene, the transition branches from dry to fleshy fruit trait, or specific sites in the sequences were tested for significantly different rates of evolution. Model 0 (M0) was used to estimate a single evolutionary rate for all genes when the clades being analyzed encompassed the entire dataset. Model 2 (M2) was used when two groups encompassing the entire data set have different rates or when two groups that are being compared do not encompass the entire data set. In the latter case, the two clades being compared were grouped together to obtain a single evolutionary rate in comparison to the rate for the remaining data (background). This single rate for the two clades grouped together was then compared to the rates for each clade separately to determine if the separate rates were significantly different from the combined rate. The test statistic, 2ΔL (twice the difference of the resulting log-likelihood values), and the degrees of freedom (df), were then used in chi-squared tests to check for statistical significance. In any comparison where the *P*-value was less than 0.05, the second hypothesis was considered to have the better fit than the first, implying there is statistical power to support that the gene clades are evolving at different rates. Since Solanaceae has a well-supported phylogeny (Olmstead et al., 2008; Särkinen et al., 2013), for PAML analyses, the branches of the gene-tree described above were adjusted to match the phylogenetic relationships of the species included in the analysis. In the *euFUL* gene groups that are evolving faster,

¹ [ftp://ftp.solgenomics.net/manuscripts/Litt_2018](http://ftp.solgenomics.net/manuscripts/Litt_2018)

² <https://www.ncbi.nlm.nih.gov/blast>

³ <https://db.cngb.org/blast4onekp/>

⁴ <https://sites.google.com/a/ualberta.ca/onekp>

⁵ <http://abacus.gene.ucl.ac.uk/software/paml.html>

sites undergoing positive selection were analyzed using mixed effects model of evolution (MEME)⁶ (Murrell et al., 2012).

The gene alignments for the *euFUL* subclades that are evolving at statistically significantly faster than the other subclades were translated using AliView (Larsson, 2014). In these protein alignments, the sites that changed from hydrophilic to hydrophobic or vice versa were identified manually. Those changes that might have been functionally deleterious versus those that might have been neutral were identified using the PROVEAN Protein tool⁷ (Choi, 2012; Choi et al., 2012; Choi and Chan, 2015).

MADS (M), intervening/interacting (I) and keratin-like (K) domains of the proteins were identified using a published MADS-box protein model (Kaufmann et al., 2005).

The structure of M, I, and K domains of tomato *FUL1* and *MBP10* were predicted using PHYRE2 server⁸ (Kelley et al., 2015).

MBP10/MBP20 Synteny and Intron Analyses

One-million-base-pair regions surrounding tomato *MBP10* and *MBP20* were analyzed for synteny using the progressive Mauve alignment tool on Geneious 10.2⁹ (Darling et al., 2010; Kearse et al., 2012).

Putative TF binding site searches for *MBP10* and *MBP20* first introns were done using PROMO 3.0¹⁰ at a maximum matrix dissimilarity rate of zero (Messeguer et al., 2002; Farré et al., 2003).

Solanaceae *euFUL* Expression Analysis

The expression patterns of *euFUL* genes were analyzed using RT-PCR data for *Solanum pimpinellifolium* organs, and transcriptome data from this study for five stages of fruit development in *S. pimpinellifolium* and tomato following stages identified by Gillaspay et al. (1993) and Tanksley (2004). Additional expression data were obtained from the eFP browser¹¹ for tomato, *S. pimpinellifolium*, potato (*S. tuberosum*) (Massa et al., 2011; Potato Genome Sequencing Consortium et al., 2011; The Tomato Genome Consortium, 2012) and from the Gene Expression Atlas¹² for *Nicotiana benthamiana* (Nakasugi et al., 2014), and other publications (Hileman et al., 2006; Burko et al., 2013).

The TF binding sites for the 2 kb and 5 kb regions upstream of the *euFUL* gene transcription start sites of tomato (GCF_000188115.4) (The Tomato Genome Consortium, 2012), potato (GCF_000226075.1) (Potato Genome Sequencing Consortium et al., 2011) and *N. sylvestris* (GCA_000393655.1) (Sierro et al., 2013) were predicted using PlantPAN 2.0¹³

(Chang et al., 2008). Due to the limitations of available contig length, the longest promoter region used for *N. sylvestris* *MBP10* was 3.3 kb.

RESULTS

Solanaceae Have Four Clades of *euFUL* Genes

Our analysis consisted of 106 sequences from 45 species in 26 genera obtained from direct amplification, transcriptomes, and online genomic databases (Table S1 in Supplementary Data Sheet S1). Of these, 64 sequences belonged to species from the Solanoideae, characterized by the derived fleshy fruit, whereas the other 42 sequences were from species with the ancestral dry-fruit trait. We designated *euFUL* genes from Convolvulaceae, the sister-group of Solanaceae, as the outgroup (Stefanović et al., 2003). For many species in the analysis, we have an incomplete set of paralogs; however, we had substantial and diverse representation from across the phylogeny, which allows us to test hypotheses regarding the evolution of this gene lineage in Solanaceae.

We used maximum likelihood methods (Garli v2.1) (Bazin et al., 2014) to reconstruct the relationships of Solanaceae *euFUL* genes (Figure 2). The resulting tree shows two major lineages of *euFUL* genes, with 80% and 100% bootstrap support, respectively, that correspond to the previously identified core eudicot *euFULI* and *euFULII* lineages (Litt and Irish, 2003; Shan et al., 2007). A Solanaceae whole-genome triplication has been proposed (The Tomato Genome Consortium, 2012; Albert and Chang, 2014; Vanneste et al., 2014; Bombarely et al., 2016), which would suggest that all Solanaceae should have three *euFULI* and three *euFULII* genes. However, others have suggested a duplication (Blanc and Wolfe, 2004; Schlueter et al., 2004; Song et al., 2012). Our data and other studies, as well as searches of the tomato genome have shown that tomato has four *euFUL* genes: two *euFULI* and two *euFULII* (Hileman et al., 2006; Bemer et al., 2012; The Tomato Genome Consortium, 2012) instead of the six predicted by a triplication. Additional genome sequencing (e.g., potato, *Capsicum annuum*) (Potato Genome Sequencing Consortium et al., 2011; Hulse-Kemp et al., 2018), transcriptome sequencing, and PCR-based analyses (this study) have also found two *euFULI* and two *euFULII* genes. This suggests the loss of one paralog from each of the *euFULI* and *euFULII* clades following a whole-genome triplication (The Tomato Genome Consortium, 2012; Albert and Chang, 2014; Vanneste et al., 2014; Bombarely et al., 2016) or, alternatively one or more duplication events (Blanc and Wolfe, 2004; Schlueter et al., 2004; Song et al., 2012).

For the purposes of this paper, we will refer to the *euFULI* and *euFULII* subclades by the name currently used for the tomato gene in each subclade (Hileman et al., 2006; Bemer et al., 2012). Thus, the two *euFULI* subclades will be referred to as the *FUL1* and *FUL2* clades, and the *euFULII* subclades will be referred to as the *MPB10* and *MBP20* subclades (Figure 2). In our gene tree, while the *FUL2*, *MBP10*, and *MBP20* clades had high bootstrap support of 83, 99 and 89%, respectively, the *FUL1*

⁶<http://datamonkey.org/meme>

⁷<http://provean.jcvi.org>

⁸<http://www.sbg.bio.ic.ac.uk/~phyre2>

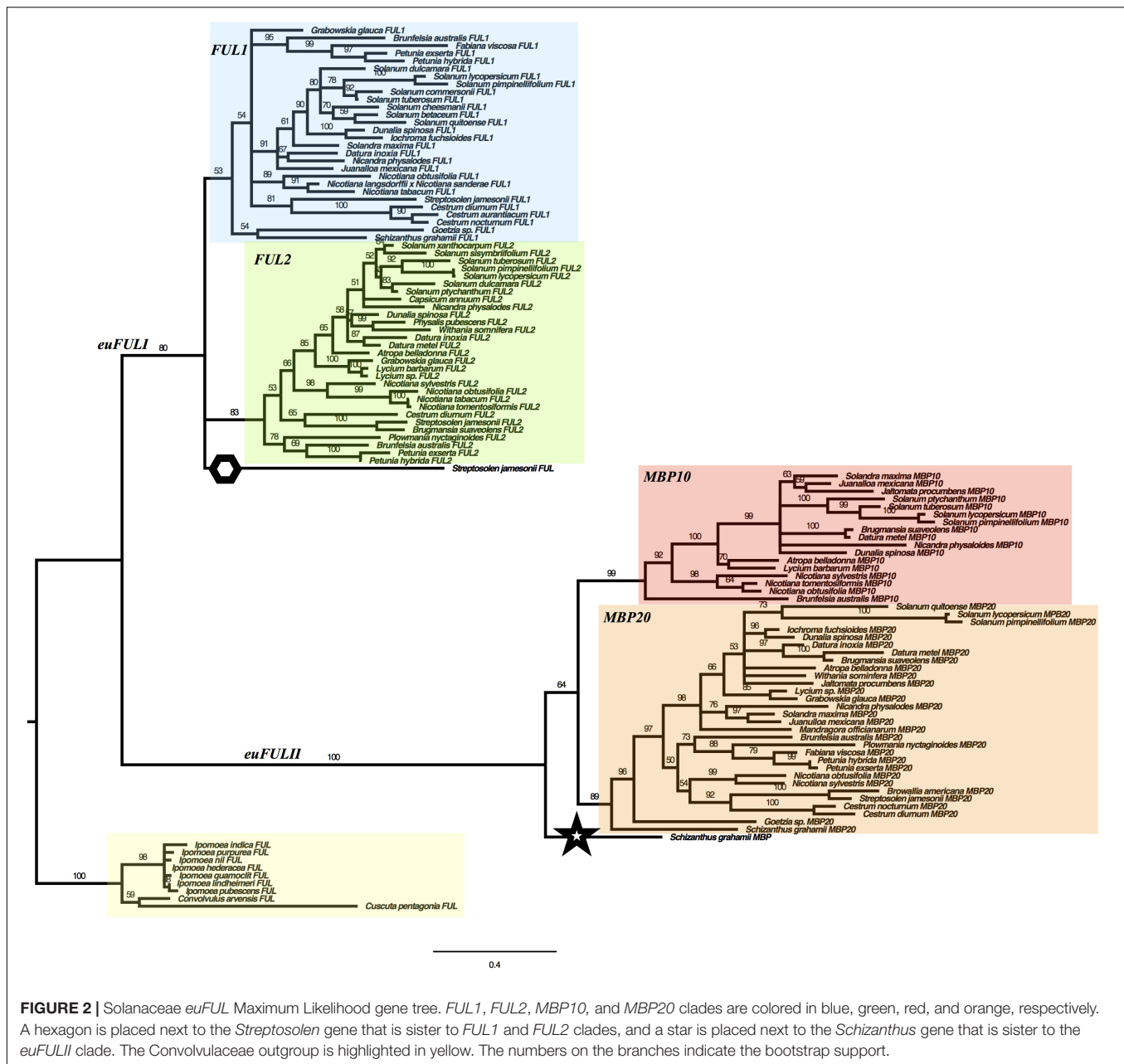
⁹<https://www.geneious.com>

¹⁰<http://algggen.lsi.upc.es/recerca/frame-recerca.html>

¹¹<http://bar.utoronto.ca>

¹²<http://benthenome.qut.edu.au/>

¹³<http://plantpan2.itps.ncku.edu.tw>



clade had only 53% support (**Figure 2**). A single gene from *Streptosolen* grouped sister to the *FUL1* and *FUL2* clades, while a gene from *Schizanthus*, one of the earliest diverging genera (Olmstead et al., 2008; Särkinen et al., 2013), grouped as sister to the *euFULII* clade. To confirm the above were not artifacts, we re-assembled the *Streptosolen* transcriptome while searching for reads supporting the gene contig, and amplified the *Schizanthus* sequence using gene-specific primers.

The presence of both *FUL1* and *FUL2* genes in species from across the phylogeny is consistent with the event that produced these two clades being part of a family-wide, whole-genome duplication or triplication (Blanc and Wolfe, 2004; Schlueter et al., 2004; Song et al., 2012; The Tomato Genome

Consortium, 2012; Albert and Chang, 2014; Vanneste et al., 2014; Bombarely et al., 2016). However, we did not find a *FUL2* ortholog in *Schizanthus*, using transcriptome data, or *Goetzia*, using PCR. These two genera are among the earliest diverging in the family (Olmstead et al., 2008; Särkinen et al., 2013), and are the earliest that we sampled. This raises the possibility that the *FUL1*/*FUL2* clades resulted from a duplication that occurred following the diversification of *Schizanthus* and *Goetzia*. In addition, although we obtained *MBP10* sequences from *Nicotiana* and most of the genera that diversified subsequently (**Figure 1** and **Figure S2** in **Supplementary Data Sheet S1**), we did not find members of the *MBP10* clade in genera that diverged prior to *Brunfelsia*. This suggests that the *MBP10* and *MBP20* subclades

were produced by a duplication that occurred later in Solanaceae diversification, after the *euFUL1* duplication and any proposed family-wide whole-genome events.

The *euFULII* Clades Are the Result of a Tandem Gene Duplication

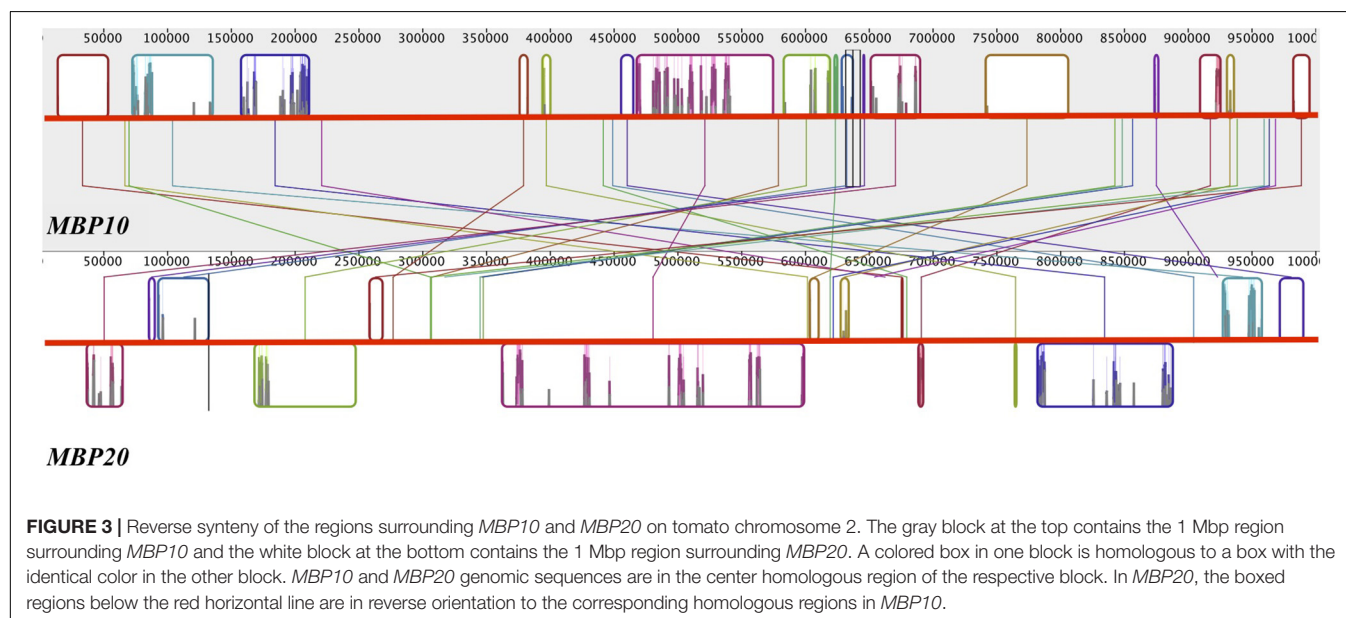
To investigate the nature of the *MBP10/MBP20* duplication, we mapped the location of the four *euFUL* paralogs to the genome of cultivated tomato. *FUL1* and *FUL2* are located on chromosomes 6 and 3, respectively, consistent with their origin from a whole genome multiplication. By contrast, *MBP10* and *MBP20* are both located on chromosome 2, about 14.3 million base pairs apart (Figure 3). The location of both *euFULII* genes on the same chromosome, and the presence of only one ortholog in early diverging species, support the hypothesis that these paralogs may be the result of a tandem gene duplication. Moreover, comparing a 1-million-base-pair region surrounding both *MBP10* and *MBP20* shows synteny, further supporting a tandem duplication (Figure 3). Annotations indicate that these syntenic zones contain 17 homologous regions. The regions that show homology are located on the opposite sides of *MBP10* and *MBP20*, suggesting an inversion of the tandemly duplicated region.

Although we recovered an *MBP10*-clade member in *Brunfelsia australis* using transcriptome analysis, we were unable to amplify this gene from leaf or floral tissue of *Fabiana* or *Plowmania*, genera that are most closely related to *Brunfelsia* (Figure 1 and Figure S2 in Supplementary Data Sheet S1). In addition, *Petunia* is also a member of the clade that includes *Brunfelsia*, and searches of the published *Petunia* genomes (Bombarely et al., 2016) also failed to turn up an *MBP10*-clade member. However, the *Brunfelsia* sequence in our analysis, obtained from transcriptome data, falls in the expected place in the phylogeny, and we confirmed the presence of *MBP10* transcript

in *Brunfelsia floribunda* floral RNA. This suggests that the *MBP10/MBP20* duplication occurred before the divergence of the *Brunfelsia/Fabiana/Petunia/Plowmania* clade but the *MBP10* paralog was lost in *Fabiana*, *Petunia* and *Plowmania*.

MBP10 Has a Short First Intron With No TF Binding Sites

A long first intron ranging from 1 to 10 kb, with multiple potential TF binding sites, is a general feature of *FUL* homologs (Table 2) (Takumi et al., 2011). By contrast, *MBP10* has a short first intron of about 80 bp in both cultivated tomato and its closest wild relative, *S. pimpinellifolium*, and about 110 bp in *Nicotiana obtusifolia* (Table 2). The expression of most *euFUL* genes is strong across nearly all vegetative and reproductive organs (Ferrándiz et al., 2000; Shchennikova et al., 2004; Kim et al., 2005; Hileman et al., 2006; Bemer et al., 2012; Pabón-Mora et al., 2012, 2013; Scorza et al., 2017); however, diverse analyses using both quantitative and non-quantitative methods indicate that *MBP10* expression is relatively weak in tomato, *S. pimpinellifolium*, and *N. obtusifolia* in most organs (Massa et al., 2011; Potato Genome Sequencing Consortium et al., 2011; The Tomato Genome Consortium, 2012; Nakasugi et al., 2014), however, some studies have suggested moderate expression in leaves (Figure S1 in Supplementary Data Sheet S1 and unpublished data). To determine if the short first intron lacks putative TF binding sites, we searched the first intron of *MBP10* and *MBP20* in tomato (Promo v3.0) (Messegueur et al., 2002; Farré et al., 2003). We found that the first intron of *MBP10* contains no putative TF binding sites, while that of *MBP20* contains 88 putative TF binding sites for eight different TFs. These TFs belong to five main families (Figure S3 in Supplementary Data Sheet S1): MYB (MYB2, C1), HSF (HSF1), Dof (Dof1, MNB1a, PBF), WRKY (SPF1) and MADS-box (SQUA). A similar situation was observed for *Nicotiana obtusifolia*, which had 133 putative binding sites in the



first intron of *MBP20* for a similar array of TFs, while *MBP10* had only four such sites. In addition, we searched the first intron of *AGL79*, the *euFULII* paralog of *FUL* in *A. thaliana*, and found 49 putative binding sites, also for similar TFs and TF families. This suggests a loss of regulatory motifs in *MBP10*.

FUL1* and *MBP10* Genes Are Evolving at a Faster Rate Than *FUL2* and *MBP20

Using the Solanaceae *euFUL* sequence data (Table S1 in Supplementary Data Sheet S1), we conducted selection pressure analyses (PAML v1.3) (Yang, 1997) to investigate if there was a shift in evolutionary rate following the *FUL1/FUL2* or *MBP10/MBP20* duplication. Selection pressure (ω) acting upon different *euFUL* gene subclades was calculated as the ratio of the rate of non-synonymous substitutions to the rate of synonymous substitutions (dN/dS) (Yang, 1997; Yang and Nielsen, 2000). An ω value of less than 1 means the coding regions are under purifying selection and that protein function is conserved. By contrast, an ω of more than 1 means that the coding regions are under diversifying selection (Yang and Nielsen, 2000). This is interpreted as allowing potential divergence in protein function (Torgerson et al., 2002; Almeida and Desalle, 2009). The nucleotide alignments we used in these analyses excluded the C-termini for all sequences except for those in the *FUL2* clade, due to the high variability of this region, which prevents reliable alignment.

Our results indicate that all Solanaceae *euFUL* gene clades are undergoing purifying selection ($\omega \leq 0.20$; Table 1 and Table S3 in Supplementary Data Sheet S1), suggesting conservation of function. The two main lineages, *euFULI* ($\omega = 0.13$) and *euFULII* ($\omega = 0.16$) are evolving at statistically indistinguishable rates. However, within the *euFULI* clade, genes of the *FUL1* clade are evolving at a significantly higher rate ($\omega = 0.17$) compared to those of the *FUL2* clade ($\omega = 0.11$). Within the *euFULII* clade, *MBP10* genes are also evolving at a significantly higher rate ($\omega = 0.19$) compared to *MBP20* ($\omega = 0.15$). Comparing each clade

against all other clades showed that *FUL2* ortholog sequences are the most conserved while *MBP10* ortholog sequences have the weakest purifying selection rates, followed by *FUL1*, implying the possibility of diversifying functions in the latter two subclades (Table 1 and Table S3 in Supplementary Data Sheet S1). None of the gene groups showed a change in evolutionary rates in comparisons between dry- and fleshy-fruited species (Table S3 in Supplementary Data Sheet S1).

Rapidly Evolving Sites Are in Regions Responsible for Protein Complex Formation

We further analyzed the sequences to identify changes at individual amino acid sites, specifically those that involved a change between polar/charged and non-polar, that might have resulted in a change in protein conformation and function and that were correlated with the change from dry to fleshy fruit. The *euFUL* genes belong to the Type II MADS-domain containing proteins, which are characterized by a MADS (M) domain, which functions in DNA binding and DNA-protein dimer specificity, an intervening/interacting (I) domain that also has a role in dimer specificity, a keratin-like (K) domain important for protein-protein interactions, and a C-terminal (C) domain, implicated in protein-multimerization, transcription activation, and additional functions (Cho et al., 1999; Heijmans et al., 2012). The C-termini were excluded from this analysis. We selected comparisons in which our results showed two gene groups evolving at significantly different rates (e.g., *FUL1* vs. *FUL2*; Table 1 and Table S3 in Supplementary Data Sheet S1). In the faster evolving group, we searched for sites in the M, I, and K regions that are undergoing diversifying selection (> 1) using mixed effects model of evolution (MEME) (see footnote 6)¹⁴ (Murrell et al., 2012). The results (Figure S4 in Supplementary Data Sheet S1) suggest that sites undergoing diversifying selection are located mainly

¹⁴<http://datamonkey.org/meme>

TABLE 1 | Evolutionary rates of *euFUL* gene clades that are evolving at statistically different rates.

Comparison	Model	ω_0	ω_1	ω_2	2 ΔL	df	P-value
<i>FUL1</i> vs. <i>FUL2</i>	M2 _A (ω_0 : background; ω_1 : <i>FUL1</i> and <i>FUL2</i>)	0.1577	0.1311	—	15.5040	1	0.0001
	M2 _B (ω_0 : background; ω_1 : <i>FUL1</i> ; ω_2 : <i>FUL2</i>)	0.1577	0.1710	0.1064			
<i>MBP10</i> vs. <i>MBP20</i>	M2 _A (ω_0 : background; ω_1 : <i>MBP10</i> and <i>MBP20</i>)	0.1311	0.1577	—	7.0291	1	0.0080
	M2 _B (ω_0 : background; ω_1 : <i>MBP10</i> ; ω_2 : <i>MBP20</i>)	0.1279	0.1939	0.1514			
<i>FUL1</i> vs. other <i>euFUL</i>	M0 (ω_0 : all branches)	0.1423	—	—	5.3906	1	0.0001
	M2 (ω_0 : <i>FUL1</i> ; ω_1 : other <i>euFUL</i>)	0.1706	0.1344	—			
<i>FUL2</i> vs. other <i>euFUL</i>	M0 (ω_0 : all branches)	0.1423	—	—	19.3663	1	0.0000
	M2 (ω_0 : <i>FUL2</i> ; ω_1 : other <i>euFUL</i>)	0.1065	0.1622	—			
<i>MBP10</i> vs. other <i>euFUL</i>	M0 (ω_0 : all branches)	0.1423	—	—	8.7258	1	0.0031
	M2 (ω_0 : <i>MBP10</i> ; ω_1 : other <i>euFUL</i>)	0.1943	0.1348	—			

2 ΔL , test statistic (two times the difference of log-likelihood values); df, degrees of freedom. When the groups being compared did not encompass the entire data set, Model 2 required a two-step analysis (see the section "Materials and Methods"). In this case subscript A and B were used to distinguish the analyses.

TABLE 2 | Approximate lengths of the first introns of several *FUL* homologs.

Gene	Length (bp)
<i>Solanum lycopersicum</i> <i>FUL1</i>	5,000
<i>S. lycopersicum</i> <i>FUL2</i>	4,400
<i>S. lycopersicum</i> <i>MBP10</i>	80
<i>S. lycopersicum</i> <i>MBP20</i>	2,500
<i>S. pimpinellifolium</i> <i>MBP10</i>	80
<i>Nicotiana obtusifolia</i> <i>FUL1</i>	5,300
<i>N. obtusifolia</i> <i>FUL2</i>	3,800
<i>N. obtusifolia</i> <i>MBP10</i>	110
<i>N. obtusifolia</i> <i>MBP20</i>	3,000
<i>Arabidopsis thaliana</i> <i>FUL</i>	900
<i>A. thaliana</i> <i>AGL79</i>	1,700
<i>Aegilops tauschii</i> <i>VRN-D1</i> (Takumi et al., 2011)	8,600

between amino acids 90 and 180 (out of ~210 amino acids in the protein). This region corresponds to the K domain (~90 to ~180 amino acids) (Kaufmann et al., 2005). In comparison, the M (~1 to ~60 amino acids) and the I domains (~60 to ~90 amino acids) had relatively few sites undergoing diversifying selection. Since these TFs function in complexes with other MADS-domain proteins as well as other proteins, novel interactions made possible by amino acid changes in this region might lead to changes in transcriptional activity.

The K domain had 14 sites undergoing diversifying selection in the *FUL1* proteins and four of those showed a change in polarity (Figure S4 in Supplementary Data Sheet S1). Of those four, a site that corresponds to the 153rd residue in the tomato protein had negatively charged glutamate (E) in most of the non-Solanoideae (mainly dry-fruited) species (11 out of 15 sequences) while all Solanoideae (mainly fleshy-fruited) species had a non-polar residue: valine (V; 13 species) or methionine (M; 1 species) (Figure S5 in Supplementary Data Sheet S1). This change was due to a single nucleotide change from an A to T in the former and G to A in the latter. All other changes in *FUL1* proteins that result in a change in charge appeared to be reversible, and none were correlated with the phylogeny nor with phenotypic changes. We used the PROVEAN tool on all four K-domain sites that showed a change in charge to predict whether these transitions were likely to be deleterious or neutral (Choi, 2012; Choi et al., 2012; Choi and Chan, 2015). Two of these sites, one with a histidine (H) to glutamine/asparagine (Q/N) shift at the 95th residue, and one with a lysine (K) to glutamine/threonine (Q/T) shift at the 157th residue (Figure S5 in Supplementary Data Sheet S1), were predicted to be functionally deleterious while the other two sites, including the 153rd residue with E to V change, were predicted to be neutral. There were five rapidly changing sites in the M domain and six sites undergoing positive selection in the I domain of *FUL1*. None of the sites in the M domain showed a change in polarity. Only one site in the I domain showed a change in polarity, but this site was predicted to be neutral functionally. *MBP10* proteins had 20 sites undergoing diversifying selection in the K domain, only 1 such site in the M domain and 3 in the I domain (Figures S4, S5 in Supplementary Data Sheet S1).

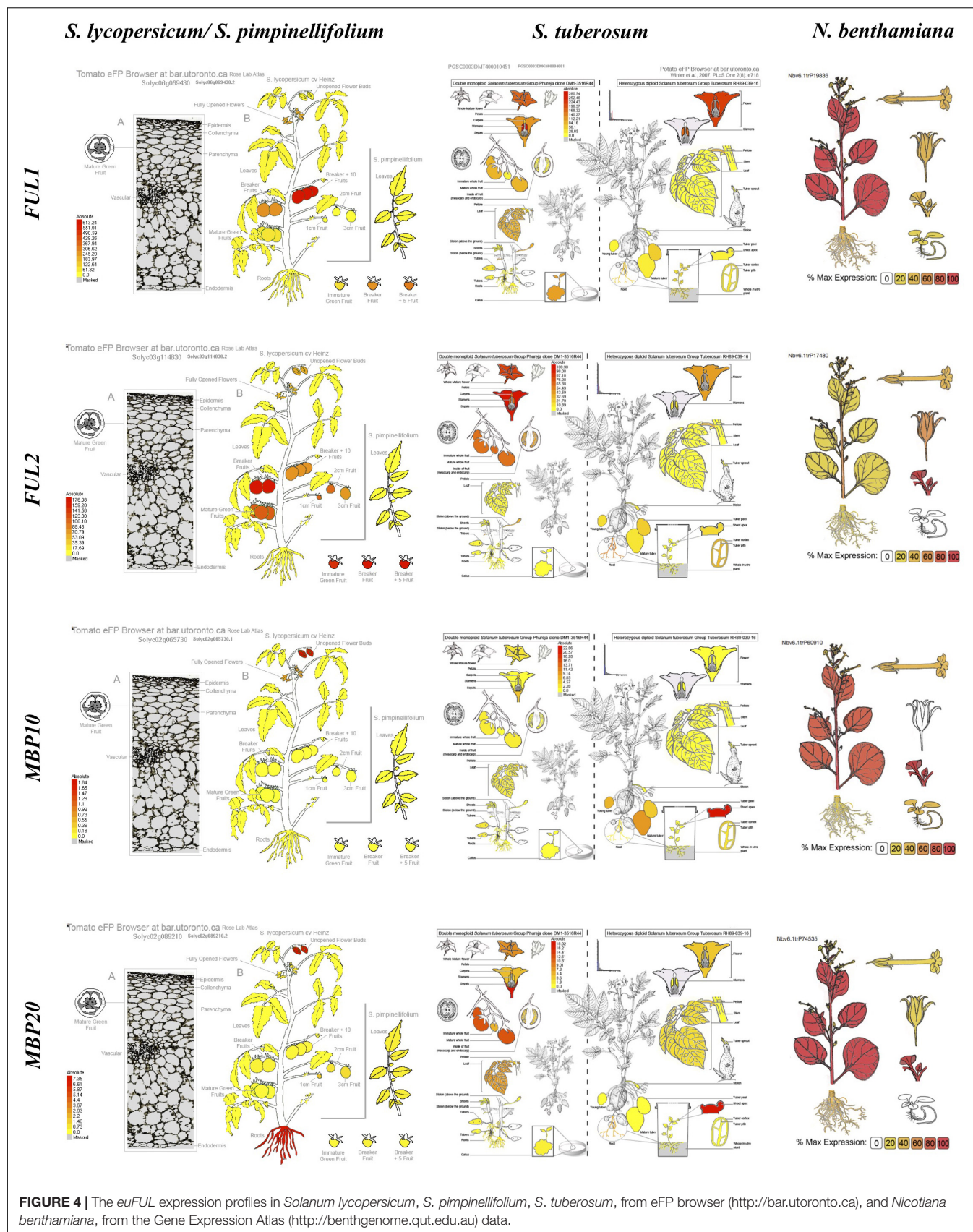
Of these, only three sites in the I domain showed a change in charge, all of which were also predicted not to have a negative effect on function.

Solanaceae *euFULI* and *euFULII* Homologs May Have Experienced Distinct Mechanisms of *Cis*-Regulatory Evolution

We compared *euFUL* expression data for the cultivated and wild tomato species, potato and *Nicotiana benthamiana* to identify any patterns that might be the result of changes in the regulatory regions following the duplications of these genes. Not all data from online sources were comparable across species, as different studies included different organs and developmental stages in their analyses, limiting cross-species comparisons. The analysis shows similar spatial expression patterns for *FUL1* and *FUL2* (Figure 4 and Figure S1 in Supplementary Data Sheet S1). These two paralogs are broadly expressed in leaves, flowers and fruits of tomato, potato, and tobacco. Although the eFP browser data (Figure 4) shows no expression for *FUL1* and *FUL2* in tomato leaves, our RT-PCR data (Figure S1 in Supplementary Data Sheet S1) and previous publications (Hileman et al., 2006; Burko et al., 2013) show expression of all four *euFUL* homologs in these organs. Both *euFULI* genes are expressed relatively weakly in the roots of tomato, potato, and tobacco (Massa et al., 2011; Potato Genome Sequencing Consortium et al., 2011; The Tomato Genome Consortium, 2012; Nakasugi et al., 2014) (Figure 4 and Figure S1 in Supplementary Data Sheet S1). Although spatial domains of expression are similar for the *euFULI* genes, they differ in temporal expression over the course of fruit developmental stages in tomato. Although both *FUL1* and *FUL2* are expressed in the fruits of all species, in tomato *FUL2* is highly expressed during the early stages of fruit development and then tapers off, whereas *FUL1* expression increases with time (Figure 4 and Figure S1 in Supplementary Data Sheet S1).

In comparison to the *euFULI* genes, the two *euFULII* paralogs show more striking differences in spatial expression at the organ level (Figure 4 and Figure S1 in Supplementary Data Sheet S1), and also between species. In all species for which expression is reported, *MBP10*, alone among the *euFUL* genes in Solanaceae, is not expressed in fruits, or is expressed at barely detectable levels. In tomato, *MBP20* is expressed strongly in roots while *MBP10* is not. By contrast, in potato tubers, *MBP10* expression is high and *MBP20* is not expressed (Figure 4). The online sources and our RT-PCR data also show subtle intra-specific differences in expression between *MBP10* and *MBP20* in flowers (Figure 4 and Figure S1 in Supplementary Data Sheet S1). In addition, our RT-PCR data show that *MBP10* is expressed relatively weakly in petals and stamens in tomato while *MBP20* is expressed throughout the flower (Figure S1 in Supplementary Data Sheet S1). However, these differences seem to be a matter of expression intensity in comparison to the more striking contrasts seen in roots, tubers, and fruits.

The types of differences in expression between *FUL1* and *FUL2* versus *MBP10* and *MBP20* might be due to differences in the regulatory environment as a result of the different ways



in which these duplicates arose. A tandem duplication and inversion may have disrupted regulatory regions in ways that would not be associated with a whole genome duplication or triplication (Tanimoto et al., 1999; Kmita et al., 2000; Vogel et al., 2009; Lupiáñez et al., 2015; Puig et al., 2015). To investigate this, we searched for putative TF binding sites in the promoter regions (2 and 5 kb upstream from the transcription start site) of *euFUL* genes in tomato, potato, and woodland tobacco to compare the differences between the pairs of paralogs (Table S4 in Supplementary Data Sheet S1). Woodland tobacco was used rather than *N. benthamiana* since relatively longer promoter sequence lengths for *euFUL* genes were available for this genome assembly (Sierro et al., 2013). Despite this, the maximum available promoter length for *NsMBP10* was about 3.3 kb. We found that the differences in types and numbers of predicted TF binding sites between *FUL1* and *FUL2* were comparable to the differences between *MBP10* and *MBP20* (Table S4 in Supplementary Data Sheet S1). Nonetheless we did find some differences that may underlie observed differences in expression between paralogs. Some of these differences were presence/absence of binding sites for a particular TF, and some were in the number and distribution of sites. Putative binding sites for AUXIN RESPONSE FACTORS (ARF) were absent from the tomato *FUL2* promoter while they were present in the promoters of all other *euFUL* genes in all species examined. Only *FUL2* in tomato, *FUL1* in potato, and *MBP10* in woodland tobacco contained binding sites for STOREKEEPER (STK). ETHYLENE INSENSITIVE 3 (EIN3) has three sites in tomato *FUL1* and five in tomato *FUL2*, but the distribution of the sites differs. In *FUL1*, there are no sites within 2 kb of the coding sequence, and three within 5 kb, whereas in *FUL2* there is one site in the 2 kb region and four in the full 5 kb region. In woodland tobacco, there are three EIN3 sites in *FUL1*, all of which are within the 2 kb region, and only one in *FUL2*, which is located between 2 and 5 kb. These types of differences may underlie observed differences in expression.

DISCUSSION

Solanaceae *euFUL* Gene Tree Shows the History of Duplications in This Lineage

In Solanaceae, there has been a major shift to fleshy fruit in the Solanoideae (Knapp, 2002). However, we do not know the molecular basis of this economically and ecologically important evolutionary event. *FUL* negatively regulates lignification in the dehiscence zone in the dry silique of *A. thaliana*, and functions in cauline leaf development, the transition to flowering and determinacy (Spence et al., 1996; Gu et al., 1998; Liljegren et al., 2000, 2004; Rajani and Sundaresan, 2001; Melzer et al., 2008). Studies of *FUL* ortholog function across the angiosperms have shown that it is labile, and orthologs have acquired diverse roles over evolutionary time. *VEGETATIVE 1* (*VEG1*), an ortholog of *FUL* in pea (*Pisum sativum*), is involved in secondary inflorescence meristem identity (Berbel et al., 2012). *AGAMOUS-like 79* (*AGL79*), the *A. thaliana euFULII* paralog of *FUL*, is mainly expressed in the root and has functions in

lateral root development and may also play a role function in leaf shape, leaf number, branching, and time to flowering (Gao et al., 2018). However, the overexpression of an *AGL79* ortholog from snapdragon (*Antirrhinum majus*) in *A. thaliana* resulted in indehiscent siliques, suggesting a role more similar to *A. thaliana FUL* (Müller et al., 2001). Evidence suggests that in tomato, one of the *AGL79* orthologs, *MBP20*, plays a role in leaf development (Burko et al., 2013). *VERNALIZATION 1* (*VRN1*) genes, which are *FUL-like* orthologs in grass species such as wheat (*Triticum* spp.) and barley (*Hordeum vulgare*), function in the vernalization response (Preston and Kellogg, 2008). Evidence to date, therefore, suggests that *euFUL* function is labile, and has changed substantially in different plant lineages during the course of angiosperm evolution. Thus it is not surprising to find a change in function of *euFUL* orthologs in Solanaceae.

There is evidence to suggest that Solanaceae *euFUL* orthologs play a role similar to that of *A. thaliana FUL* in the development of dry dehiscent fruits (Smykal et al., 2007). However, studies suggest that in the fleshy fruit of Solanoideae, *FUL* orthologs play roles in pigmentation as well as ethylene response, cell wall modification, glutamic acid degradation, volatile production, and pericarp and cuticle thickness (Bemer et al., 2012; Shima et al., 2014; Wang et al., 2014). To determine if we could identify changes in *euFUL* sequences or selection that might shed light on this change in function, we analyzed *euFUL* gene evolution in Solanaceae.

We performed a maximum likelihood phylogenetic analysis (Garli v2.1) (Bazin et al., 2014) on a data set that consisted of 106 Solanaceae members of the *euFUL* gene lineage (Litt and Irish, 2003; Shan et al., 2007), which we obtained through amplification and sequencing (37 sequences), generating transcriptome sequence data (29 sequences), or mining databases (40 sequences). As outgroup we used 10 *euFUL* genes from Convolvulaceae, the sister family to Solanaceae (Figure 2 and Supplementary Table S1) (Stefanović et al., 2003). The resulting tree shows the two major clades of core-eudicot *euFUL* genes, the *euFULI* and *euFULII* lineages (Litt and Irish, 2003; Shan et al., 2007). Within each of these clades there is evidence of a Solanaceae-specific duplication, resulting in two subclades in each lineage. Within each subclade, the order of branches correlates well with the topology of the Solanaceae phylogeny (Olmstead et al., 2008; Särkinen et al., 2013); discrepancies at the genus level are likely due to the short length of some sequences and sequence divergence in some taxa. Each of the subclades includes orthologs from both fleshy- and dry-fruited species, indicating that the subclade duplications preceded the origin of fleshy fruit.

Although duplications in these genes are common (Litt and Irish, 2003; Preston and Kellogg, 2007; Pabón-Mora et al., 2013), we did not find significant evidence of taxon-specific duplications. We did, however, find two genes that did not fall into a specific subclade. A third *Streptosolen* gene grouped sister to the rest of the *euFULI* clade (76% identity among the three *Streptosolen* genes), potentially the result of a taxon-specific duplication followed by sequence divergence. In addition, a *Schizanthus* gene grouped sister to the *euFULII* clade (77%

pairwise identity with *Schizanthus MBP20*). This may also be a divergent genus-specific paralog, but since *Schizanthus* is one of the earliest diverging genera (Olmstead et al., 2008; Särkinen et al., 2013), it is also possible this gene might be a remaining paralog from the reported whole genome duplication/triplication that occurred early in Solanaceae diversification (Blanc and Wolfe, 2004; Schlueter et al., 2004; Song et al., 2012; The Tomato Genome Consortium, 2012; Albert and Chang, 2014; Bombarely et al., 2016). Examination of sequences showed that these are not likely to be splice isoforms. We also found potential evidence of loss – not every Solanaceae species we studied had a copy of each *euFUL* gene. We did not, for example, find *FUL2* genes in *Iochroma*, *Fabiana*, *Solandra*, *Juanulloa*, *Schizanthus*, or *Goetzia*, even though these all had genes in the *FUL1* clade (see **Table S1** in **Supplementary Data Sheet S1** for a complete list). However, although this may represent paralog loss, it is possible we did not recover all gene copies due to PCR primer mismatches, low expression levels, or the absence of transcript in the sampled tissue.

In addition to the major shift to fleshy fruit in the Solanoideae subfamily, fleshy fruits have independently evolved in *Cestrum* and *Duboisia*, and there has also been a reversal to a dry fruit in *Datura* (Knapp, 2002). Our analysis does not include genes from *Duboisia*, but the *euFUL* genes from *Cestrum* and *Datura* grouped in positions in the tree that were expected based on their phylogenetic position, and did not show any notable differences in sequence from the *euFUL* genes of their close relatives.

***euFULI* and *euFULII* Clade Duplicates Have Experienced Different Levels of Purifying Selection**

We compared dN/dS ratios between and among Solanaceae *euFULI* and *euFULII* lineages, as well as between sequences before and after the transition to fleshy fruit, to investigate if any changes in selection might be correlated with sequence diversification. All ω values from our analyses are closer to 0 than to 1 (**Table 1** and **Table S3** in **Supplementary Data Sheet S1**), which indicates that all *euFUL* gene clades are under strong purifying selection (Yang, 1997; Yang and Nielsen, 2000; Torgerson et al., 2002; Almeida and Desalle, 2009). Studies suggest that this is the norm for most protein coding genes, and that under such stringent evolutionary constraints, slight differences in evolutionary rates may result in functional diversification (Yu et al., 2017). Our data show a weakening of purifying selection in *FUL1* genes relative to *FUL2* genes ($\omega = 0.17$ vs. 0.11 , $p < 0.0005$) and in *MBP10* genes relative to *MBP20* genes ($\omega = 0.19$ vs. 0.15 , $p < 0.01$). Immediately after the *euFULI* duplication, the *FUL1* and *FUL2* lineage genes would have been fully redundant, which might have allowed the reduction in purifying selection on the *FUL1* genes resulting in potential functional divergence. Similarly, the duplication that resulted in the two *euFULII* gene clades would have resulted in redundancy in the *MBP10* and *MBP20* lineages, possibly allowing the more rapid diversification of *MBP10* genes.

Although studies indicate that the *euFULI* genes of tomato have novel functions compared to those in dry fruit (Gu et al.,

1998; Smykal et al., 2007; Bemer et al., 2012; Shima et al., 2014; Wang et al., 2014), it remains unclear whether the new functions are the result of changes in coding sequences, regulatory regions, or downstream gene targets. Our analysis shows that *euFUL* genes in both dry- and fleshy-fruited species are evolving at similar rates (**Table S3** in **Supplementary Data Sheet S1**). This suggests conservation of the coding sequences in both fleshy- and dry-fruited species despite the central roles in the development of these distinct fruit morphologies.

Sixty-four of the sequences in our analysis were from fleshy-fruited species whereas only 42 were from dry-fruited species. Although, we had broad representation across the dry grade, it is possible with additional representation from dry fruited species, more evolutionary patterns would be revealed (Anisimova et al., 2001; Domazet-Loso, 2003; Nielsen et al., 2005).

FUL1 and MBP10 Proteins Show Amino Acid Changes in Conserved Functional Domains

An analysis of selection across an entire sequence may indicate different types of selection for the whole gene, but this overlooks the fact that key residues may be undergoing rapid evolution that may result in functional changes (Ota and Nei, 1994; Nei et al., 1997; Yang and Bielawski, 2000; Piontkivska et al., 2002; Martinez-Castilla and Alvarez-Buylla, 2003; Jeffares et al., 2006). Other empirical studies have further described functional changes due to a change in a single amino acid residue (Ingram, 1957; Hanzawa et al., 2005; Hichri et al., 2011; Zhao et al., 2012; Fourquin et al., 2013; Dai et al., 2016; Sakuma et al., 2017) specifically associated with changes in polarity (Schröfelbauer et al., 2004; Hoekstra et al., 2006) or conformation (Aseev et al., 2012). Studies in *A. thaliana*, show that a single amino acid mutation in GLABRA1 (GL1) results in the inhibition of trichome formation (Dai et al., 2016) and a change of a single residue is sufficient to convert the function of TERMINAL FLOWER 1 (TFL1), which inhibits flower formation, to that of the closely related FLOWERING LOCUS T (FT), which promotes flowering (Hanzawa et al., 2005). Three-dimensional modeling has also shown that a single amino acid change in a highly conserved domains may lead to changes in protein–protein interactions (Teng et al., 2009; David et al., 2012; Li et al., 2014). We searched for individual sites in the predicted amino acid sequences that showed evidence of positive selection within the gene groups that, although under purifying selection, were found to have statistically significantly accelerated evolutionary rates (i.e., the *FUL1* and *MBP10* clades) to determine if any amino acid changes at these sites had the potential to result in a change in protein function.

Our findings show that more residues are rapidly changing in the K domain compared to the M and I domains (**Figure S4** in **Supplementary Data Sheet S1**). The K domain is predicted to have an α -helix structure that facilitates protein–protein interactions (**Figure S5** in **Supplementary Data Sheet S1**) (Yang et al., 2003a,b; Kaufmann et al., 2005; Immink et al., 2010). The α -helix structure depends on conserved hydrophobic residues spaced through the domain (Eisenberg et al., 1982). Therefore,

changes to protein residues that alter charge and/or conformation in this region can lead to changes in such interactions. Most of the rapidly evolving sites did not show an amino acid change specifically associated with the shift to fleshy fruit, but rather showed changes and reversals over the course of gene evolution. Interestingly, in the *FUL1* proteins, we found one site in the K domain, corresponding to the 153rd residue in the tomato protein (Slugina et al., 2018), at which 11 out of 15 sequences from dry-fruited species have a negatively charged glutamate (E) residue. In comparison, 100% of the fleshy clade contains a non-polar residue: valine (V) (13 species) or methionine (1 species). However, since the remaining four *FUL1* sequences from dry-fruited species have non-polar glutamine (Q) or V at this site, the change from charged to non-polar is not associated with the shift to fleshy fruit. In addition, a PROVEAN analysis predicted the changes at this site to be neutral with regards to function.

Two other sites in the *FUL1* K domain show changes that are predicted to have functionally deleterious consequences according to our PROVEAN analysis (Choi, 2012; Choi et al., 2012; Choi and Chan, 2015). These include a charged histidine (H) to a non-polar glutamine/asparagine (Q/N) transition at the 95th residue and a charged lysine (K) to non-polar glutamine/threonine (Q/T) transition at the 157th residue (Figure S5 in Supplementary Data Sheet S1). Polar residues are important for protein–protein interactions of the K domain α -helix (Sheinerman et al., 2000; Curran and Engelman, 2003; Ma et al., 2003; Zhou et al., 2018) and changes might disrupt interactions with other proteins (Liu et al., 2014). However, since these changes are not correlated with the fruit type, it seems unlikely that any alteration to protein function affects fruit morphology. It is also plausible that any negative effect at these sites is masked by the *FUL2* paralog, which is likely to be functionally redundant (Bemer et al., 2012; Wang et al., 2014). This is consistent with *FUL1* evolving relatively faster (Table 1 and Table S3 in Supplementary Data Sheet S1), thus enabling divergence compared to *FUL2*, which appears to be more highly functionally conserved based on stricter sequence conservation.

None of the sites undergoing positive change in the K domain of *MBP10* showed a change in charge, suggesting these changes are not likely to affect protein function. We also observed residues in the M domain that are under diversifying selection in both the *FUL1* and *MBP10* clades. These residues are located not in the α -helix region that directly binds to DNA, but in the β -sheet region of the MADS domain (Figure S4 in Supplementary Data Sheet S1) (Immink et al., 2010). β -sheets are important for protein arrangement in three dimensional space. Therefore, any changes in this region might change protein conformation, influencing DNA binding of the α -helix as well as the ability of the euFUL proteins to form higher order complexes (Pellegrini et al., 1995). However, these shifts were reversible, with no phylogenetic pattern or change in charge, and there was no correlation with the fruit type. Therefore it is unlikely that these shifts have significant functional impact.

A previous report that investigated the evolution of MADS-box genes in *A. thaliana* also found rapidly evolving sites in the M and K domains of Type II MADS-box proteins, which might have been involved in the functional diversification of this

group, but did not report changes in the I domain (Martinez-Castilla and Alvarez-Buylla, 2003). Residues in this domain that are directly involved in forming an α -helix structure are expected to be highly conserved, whereas the remaining residues may not be under such constraints (Yang et al., 2003a,b; Kaufmann et al., 2005). We found residues in the conserved region of the I domain that are undergoing diversifying selection in both *FUL1* and *MBP10* clades. Of these, one site in *FUL1* and three sites in *MBP10* had undergone changes in charge but none were predicted to negatively affect the function (Figures S4, S5 in Supplementary Data Sheet S1). In addition, as with the sites in the M and K domains, none of these was correlated with the Solanaceae phylogeny or changes in fruit morphology. It has been reported that higher rates of substitution in lineages that show weakened purifying selection or even diversifying selection may be occurring at residues of minimal functional importance (Jacobsen et al., 2016). This might explain the apparent ease of reversibility and lack of phylogenetic signal among the rapidly changing sites we observed.

The *MBP10* and *MBP20* Clades Are the Result of a Tandem Duplication Event

The *FUL1* and *FUL2* genes of tomato are located on different chromosomes (6 and 3, respectively), which is consistent with the proposed Solanaceae whole genome duplication (Blanc and Wolfe, 2004; Schlueter et al., 2004; Song et al., 2012) or triplication (The Tomato Genome Consortium, 2012; Albert and Chang, 2014; Vanneste et al., 2014; Bombarely et al., 2016) followed by loss of one paralog. The lack of a *FUL2* ortholog in our dataset from *Goetzia* or *Schizanthus* (Figure 2), the two earliest diverging genera that we included in our analyses, raises the possibility that the *FUL1/FUL2* clades originated via a duplication that occurred after the diversification of these genera, and not as a result of a whole genome event that preceded the diversification of the family. Whole genome sequences from multiple early diverging lineages will be needed to determine the timing and nature of these early events.

We did not recover an *MBP10* ortholog from any of the genera that diverged prior to *Brunfelsia* (Figure 2 and Figure S2 in Supplementary Data Sheet S1). Our investigation revealed that in tomato, both *MBP10* and *MBP20* are located on chromosome 2, about 14.3 million base pairs apart. The 1 million base-pair region surrounding each gene shows synteny, but the order of the homologous regions is reversed (Figure 3). Together, this suggests that the *MBP10/MBP20* clades are the result of a tandem duplication accompanied by an inversion (Purugganan et al., 1995; Vision et al., 2000; Achaz et al., 2001; Prince and Pickett, 2002). Supporting this, a previous report that investigated genomic duplication events in tomato also found evidence for large-scale intra-chromosomal duplications in chromosome 2 (Song et al., 2012). Although the authors suggest this event was concurrent with a whole genome duplication at the origin of the family, they give a large window, 36–82 million years ago (MYA), for the timing of this event. The stem age of the family is predicted to be approximately 49 MYA (Särkinen et al., 2013), indicating that this duplication might have happened later in Solanaceae diversification. Our data suggest that this duplication event is

independent of the reported whole genome events, occurring prior to the diversification of the *Brunfelsia* clade but after the event that produced the *FUL1* and *FUL2* clades (**Figure 2** and **Figure S2** in **Supplementary Data Sheet S1**).

The expected topology for the *euFULII* clade, based on a duplication prior to the divergence of the *Brunfelsia* clade, would be a paraphyletic grade of pre-duplication *euFULII* genes, from species that diversified prior to *Brunfelsia*, and nested *MBP10* and *MBP20* clades that would include post-duplication genes from all species that diversified subsequent to the duplication. However, in our tree, the pre-duplication genes do not form such a basal grade (**Figure 2**). Rather, they form a clade with the post-duplication *MBP20* genes. The results of our PAML analyses indicate that the *MBP20*-clade genes show less sequence divergence than *MBP10* genes; this higher degree of similarity among pre-duplication sequences and post-duplication *MBP20* genes may underlie their grouping into one clade (Pegueroles et al., 2013).

Our results indicate that the *euFULII* duplication occurred prior to the origin of the clade containing *Brunfelsia*. We would therefore expect to find both an *MBP10* and an *MBP20* in all species of that clade. However, we did not find an *MBP10* ortholog in members of this clade other than *Brunfelsia*. *MBP10* appears to have been lost from the genome of *Petunia*, based on analyses of multiple fully sequenced genomes (Bombarely et al., 2016), and potentially from *Plowmania* and *Fabiana*. We were able to recover *MBP10* orthologs from *Nicotiana* and most other later-diverging genera. However, our analysis includes fewer species from the dry grade of the Solanaceae phylogeny than the fleshy-fruited Solanoideae clade (17 out of 45) and even fewer species that diverged prior to *Brunfelsia* (7). In the *MBP10* clade in particular, our analysis includes 13 orthologs from species in the fleshy-fruited clade but just four from the dry-fruited species, and our analysis only includes sequence data from four genera that diverged prior to the origin of the *Brunfelsia* clade (*Streptosolen*, *Cestrum*, *Goetzia*, *Schizanthus*) (**Figures 1, 2**). Thus there may be genera that originated prior to *Brunfelsia* that contain *MBP10* that our sampling did not include. Floral and fruit transcriptomes, which provided *MBP10* orthologs from later diverging species, yielded no *MBP10* sequences from *Cestrum* and *Schizanthus*; nonetheless, whole genome sequences of early diverging species are needed to determine the timing of the *MBP10/MBP20* duplication.

***euFULII* Expression Divergence May Be Associated With *Cis*-Regulatory Re-coupling**

Our analysis of Solanaceae *euFUL* homologs show that *FUL1* and *FUL2* are broadly expressed in leaves, flowers, and fruit (**Figure 4** and **Figure S1** in **Supplementary Data Sheet S1**). This overall similarity in expression may indicate a conservation of *cis*-regulatory elements in gene copies following duplication (Haberer et al., 2004). Supporting this, our investigation into the number of putative TF binding sites in the promoter region of *euFULI* homologs did not reveal statistically significant differences (**Table S4** in **Supplementary Data Sheet S1**). In

tomato fruit development, *FUL1* expression increases with time, whereas *FUL2* expression reaches a maximum at early stages and then decreases over later stages (**Figure 4** and **Figure S1** in **Supplementary Data Sheet S1**). This variation in expression associated with the developmental stages might be due to changes in *cis*-elements as a result of the accumulation of random mutations over time (Force et al., 1999; Haberer et al., 2004).

Our analysis did find differences in the number and location of predicted binding sites for specific TFs or families, for instance for ARF, STK, and EIN3 TFs, which may account for the types of differences in expression seen between *euFUL* paralogs. The 5 kb region upstream of the *FUL1* transcription start site in tomato contains three putative ARF binding sites but the corresponding region of *FUL2* in tomato contains no such motifs (**Table S4** in **Supplementary Data Sheet S1**). ARF TFs, important in tomato fruit development, are activated in response to auxin and may upregulate or repress downstream genes (de Jong et al., 2010, 2015; Liu et al., 2018); the absence of binding sites from the *FUL2* promoter is the type of factor that might underlie differences in expression observed between *FUL1* and *FUL2*. Predicted STK binding sites are only found in the promoters of potato *FUL1*, tomato *FUL2* and woodland tobacco *MBP10*. STK and STK-like proteins appear to function in storage protein synthesis, glucose reception, and vegetative and reproductive development (Zourelidou et al., 2002; Curaba et al., 2003; Bömer et al., 2011; Chung et al., 2016; Nietzsche et al., 2017). Meanwhile, the 2 kb upstream region of *FUL2* contains a putative site for EIN3. This protein is involved in the development of tomato in response to ripening-associated ethylene production (Tieman et al., 2001). No such motifs are found in the corresponding region of *FUL1*. In contrast, the 2–5 kb region in *FUL2* contains four putative sites for EIN3 while the corresponding region in *FUL1* contains three such sites (**Table S4** in **Supplementary Data Sheet S1**). Such variation in number and location of TF binding sites has been shown to be associated with the temporal differences in gene expression (Lebrecht et al., 2005; Liu et al., 2006; Giorgetti et al., 2010; Guertin and Lis, 2010; White et al., 2013; Ezer et al., 2014; Levo et al., 2015; Payne and Wagner, 2015).

Whereas the *euFULI* members largely overlap in spatial expression with some variation associated with developmental stages, the *euFULII* homologs show less consistent spatial expression patterns. Only *MBP20* is expressed in tomato roots and potato fruit while only *MBP10* is expressed in potato tubers (**Figure 4**). However, these “on” or “off” expression patterns cannot be explained by the presence or absence of any putative TF binding sites (**Table S4** in **Supplementary Data Sheet S1**). These two paralogs, which appear to be the result of a tandem duplication and inversion, are located approximately 14.3 Mbp apart (**Figure 3**) on chromosome 2. Although gene clusters resulting from tandem duplications are often coexpressed, this is not the case when there are large physical distances between the genes (Lercher et al., 2003). An investigation into the expression of human transgenes in mice also found changes in expression as a consequence of an inversion, possibly through disrupting enhancer activity or changes to chromatin structure (Tanimoto et al., 1999; Vogel et al., 2009; Puig et al., 2015). Chromosomal rearrangements

such as inversions may also result in novel connections between coding regions and other promoters or long distance regulatory motifs while disrupting the original regulatory mechanisms (Kmita et al., 2000; Lupiáñez et al., 2015). This sort of re-coupling of one of the two paralogs might lead to the types of contrasting expression patterns observed for *MBP10* and *MBP20*. However, the expression patterns are not consistent across species (**Figure 4** and **Figure S1** in **Supplementary Data Sheet S1**) and this might be due to additional changes following the inversion (Cosner et al., 1997; Lupski, 1998; Haberle et al., 2008; Chiang et al., 2012). An in-depth analysis of the entire loci and their genomic environment for all paralogs in multiple species would be necessary to determine if the tandem duplication and inversion are associated with changes in proximity to heterochromatin, additional rearrangements, or other phenomena that might have altered gene expression.

***MBP10* Shows Signs of Pseudogenization**

The first intron of some MADS-box genes contains *cis*-elements important for the regulation of expression (Gazzani et al., 2003; Michaels et al., 2003; Schauer et al., 2009). Studies have found that deletions in the first intron of a *FUL*-like gene in *Aegilops tauschii* alters expression and results in the loss of the vernalization requirement (Fu et al., 2005; Takumi et al., 2011). Consistent with this, the first introns of angiosperm *euFUL* orthologs are generally in the range of 1–10 kb (**Table 2**) (Takumi et al., 2011). In contrast, tomato *MBP10* has a short first intron of 80 bp. We compared the putative TF binding sites in the first introns of *MBP10* and *MBP20* in tomato to characterize potential loss of such sites, which might suggest reduced gene regulation. The first intron of *MBP10* is predicted to have no TF binding sites, while the first intron of *MBP20* is predicted to contain 88 TF binding sites (**Figure S3** in **Supplementary Data Sheet S1**). These included binding sites for MYB, HSF, Dof, WRKY, and MADS-box TFs. Specific TFs predicted to bind to these sites include MYB2 and C1 (MYB), which play roles in anthocyanin accumulation and lignin biosynthesis, PBF (Dof), which plays a role in endosperm storage protein accumulation, and SPF1 (WRKY), thought to function in fruit ripening (Bovy et al., 2002; Fei et al., 2004; Hwang et al., 2004; Lee et al., 2010; Xu et al., 2014; Jun et al., 2015). A similar pattern was found in analysis of the first intron of *MBP10* in *Nicotiana obtusifolia*, which is 110 bp (**Table 2**). This analysis found three putative TF binding sites for MYB2 and one for PBF. By contrast, the first intron of *N. obtusifolia MBP20* is predicted to have 133 TF binding sites and include a repertoire similar to those found for tomato *MBP20*. To determine whether the difference in TF binding site number between the paralogs represented a gain of sites in the *MBP20* genes or a loss in the *MBP10* genes, we also searched for TF binding sites in the first intron of *AGL79*, the single *euFULII* ortholog in *A. thaliana* (Gao et al., 2018). We found that it contains 49 predicted TF binding sites for five different TFs in four families: MYB (MYB2, GAMYB), HSF (HSF1), WRKY (SPF1), and GT-box (GT-1). Although

this number is substantially smaller than the number of sites predicted in the first introns of the Solanaceae *MBP20* genes, the results suggest that there has been a loss of TF binding sites in *MBP10*.

Core-eudicot *euFUL* and basal-eudicot *FUL*-like genes frequently have broad expression patterns and are generally expressed in fruit (Ferrándiz et al., 2000; Shchennikova et al., 2004; Kim et al., 2005; Hileman et al., 2006; Bemer et al., 2012; Pabón-Mora et al., 2012, 2013; Scorza et al., 2017). Therefore, the absence or extremely weak expression of *MBP10* in fruits of all species, and its weak expression in most organs of tomato and potato is notable (**Figure 4** and **Figure S1** in **Supplementary Data Sheet S1**). This relatively weak expression may at least in part be due to the loss of TF binding sites in the first intron and suggests a potentially reduced role in regulating fruit-related developmental processes. Importantly, the loss of putative TF binding sites and low expression, combined with the faster evolutionary rate, suggest *MBP10* might be in the process of becoming a pseudogene. Further support for this hypothesis comes from an examination of the *MBP10* sequences, which suggests that at least two of the sequences in our study (from *N. sylvestris* and *Dunalia spinosa*) show a frameshift that would result in an premature stop codon.

CONCLUSION

Our results suggest that there was a weakening in purifying selection following the *euFUL* gene duplications in Solanaceae, resulting in coding sequence diversification in *FUL1* and *MBP10* clades relative to *FUL2* and *MBP20*. Expression of the *euFULI* genes is broad, while the *euFULII* genes have contrasting patterns at the organ level, potentially resulting from *cis*-regulatory changes associated with the inversion event. We also found evidence to suggest that the *MBP10* clade is becoming a pseudogene. Although at least some clades of Solanaceae *euFUL* genes took on new functions associated with the development of fleshy fruit we did not find any amino acid shifts that were correlated with the change in fruit type. It is also possible that the novel functions are a consequence of downstream changes, perhaps as the result of changes in binding partners or targets. Therefore, the mechanism underlying the shift in *euFUL* function from dry to fleshy fruit in Solanaceae awaits additional analyses.

AUTHOR CONTRIBUTIONS

AL designed and supervised research, and assisted in writing the paper. DM contributed to the design of the study, generated *Cestrum diurnum*, *C. nocturnum*, and *Schizanthus grahamii* transcriptome libraries, retrieved sequences from PCR-based methods and database mining, analyzed the data, and wrote the paper. CE assisted with PAML analysis, contributed suggestions for analyses, and made suggestions

on the paper. AR generated *Dunalia spinosa*, *Fabiana viscosa*, *Grabowskia glauca*, and *Salpiglossis sinuata* transcriptome libraries, contributed suggestions for analyses, contributed in recording the associated protocols, and commented on the paper. JM retrieved sequences from PCR-based methods. MS generated the *Nicotiana obtusifolia* transcriptome libraries and additional sequences using PCR-based methods. NP-M generated *Brunfelsia australis* and *Streptosolen jamesonii* transcriptome libraries, contributed suggestions for analyses, and made suggestions on this paper.

FUNDING

This research was funded by the National Science Foundation (IOS 1456109). DM was partly supported by UCR Graduate Assistance in Areas of National Need (GAANN) and Graduate Research Mentoring Program (GRMP) fellowships. NP-M acknowledges grants by COLCIENCIAS (111565842812), the Convocatoria Programáticas 2017–2018, and the Estrategia de Sostenibilidad 2018–2019, in the

Universidad de Antioquia given to the group Evo-Devo en Plantas.

ACKNOWLEDGMENTS

We are grateful for the help with transcriptome library compilation and analyses by Glenn Hicks, John Weger, Holly Eckelhoefer, and Clay Clark at the UCR IIGB Core Facility, the advice on the gene tree generation by John Heraty, feedback by Elizabeth McCarthy, Jacob Landis, and Patricia Springer and Jaimie Van Norman lab members, and greenhouse and lab support by Arman Baghaei, Alan Le, and Victor Herrera.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.00043/full#supplementary-material>

REFERENCES

- Achaz, G., Netter, P., and Coissac, E. (2001). Study of intrachromosomal duplications among the eukaryote genomes. *Mol. Biol. Evol.* 18, 2280–2288. doi: 10.1093/oxfordjournals.molbev.a003774
- Albert, V. A., and Chang, T.-H. (2014). Evolution of a hot genome. *Proc. Natl. Acad. Sci. U.S.A.* 111, 5069–5070. doi: 10.1073/pnas.1402378111
- Almeida, F. C., and Desalle, R. (2009). Orthology, function and evolution of accessory gland proteins in the *Drosophila repleta* group. *Genetics* 181, 235–245. doi: 10.1534/genetics.108.096263
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Anisimova, M., Bielawski, J. P., and Yang, Z. (2001). Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol. Biol. Evol.* 18, 1585–1592. doi: 10.1093/oxfordjournals.molbev.a003945
- Aseev, L. V., Chugunov, A. O., Efremov, R. G., and Boni, I. V. (2012). A single missense mutation in a coiled-coil domain of *Escherichia coli* ribosomal protein S2 confers a thermosensitive phenotype that can be suppressed by ribosomal protein S1. *J. Bacteriol.* 195, 95–104. doi: 10.1128/JB.01305-12
- Bazin, A. L., Zwickl, D. J., and Cummings, M. P. (2014). A gateway for phylogenetic analysis powered by grid computing featuring GARLI 2.0. *Syst. Biol.* 63, 812–818. doi: 10.1093/sysbio/syu031
- Bemer, M., Karlova, R., Ballester, A. R., Tikunov, Y. M., Bovy, A. G., Wolters-Arts, M., et al. (2012). The tomato *FRUITFULL* homologs TDR4/FUL1 and MBP7/FUL2 regulate ethylene-independent aspects of fruit ripening. *Plant Cell* 24, 4437–4451. doi: 10.1105/tpc.112.103283
- Berbel, A., Ferrándiz, C., Hecht, V., Dalmais, M., Lund, O. S., Susmilch, F. C., et al. (2012). *VEGETATIVE1* is essential for development of the compound inflorescence in pea. *Nat. Commun.* 3:797. doi: 10.1038/ncomms1801
- Blanc, G., and Wolfe, K. H. (2004). Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* 16, 1667–1678. doi: 10.1105/tpc.021345
- BLAST® Command Line Applications User Manual (2008). *BLAST® Command Line Applications User Manual*. Bethesda, MD: National Center for Biotechnology Information.
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Bolmgren, K., and Eriksson, O. (2010). Seed mass and the evolution of fleshy fruits in angiosperms. *Oikos* 119, 707–718. doi: 10.1111/j.1600-0706.2009.17944.x
- Bombarely, A., Moser, M., Amrad, A., Bao, M., Bapaume, L., Barry, C. S., et al. (2016). Insight into the evolution of the Solanaceae from the parental genomes of *Petunia hybrida*. *Nat. Plants* 2:16074. doi: 10.1038/nplants.2016.74
- Bömer, M., Uhrig, J., Jach, G., and Müller, K. (2011). Increased vegetative development and sturdiness of storekeeper-transgenic tobacco. *Open Life Sci.* 6, 342–351. doi: 10.2478/s11535-011-0009-9
- Bovy, A., de Vos, R., Kemper, M., Schijlen, E., Almenar Pertejo, M., Muir, S., et al. (2002). High-flavonol tomatoes resulting from the heterologous expression of the maize transcription factor genes LC and C1. *Plant Cell* 14, 2509–2526. doi: 10.1105/tpc.004218
- Burko, Y., Shleizer-Burko, S., Yanai, O., Schwartz, I., Zelnik, I. D., Jacob-Hirsch, J., et al. (2013). A role for *APETALA1*/fruitfull transcription factors in tomato leaf development. *Plant Cell* 25, 2070–2083. doi: 10.1105/tpc.113.113035
- Chang, W.-C., Lee, T.-Y., Huang, H.-D., Huang, H.-Y., and Pan, R.-L. (2008). PlantPAN: plant promoter analysis navigator, for identifying combinatorial cis-regulatory elements with distance constraint in plant gene groups. *BMC Genomics* 9:561. doi: 10.1186/1471-2164-9-561
- Chiang, C., Jacobsen, J. C., Ernst, C., Hanscom, C., Heilbut, A., Blumenthal, I., et al. (2012). Complex reorganization and predominant non-homologous repair following chromosomal breakage in karyotypically balanced germline rearrangements and transgenic integration. *Nat. Genet.* 44, 390–397. doi: 10.1038/ng.2202
- Cho, S., Jang, S., Chae, S., Chung, K. M., Moon, Y. H., An, G., et al. (1999). Analysis of the C-terminal region of *Arabidopsis thaliana* *APETALA1* as a transcription activation domain. *Plant Mol. Biol.* 40, 419–429. doi: 10.1023/A:100627312
- Choi, Y. (2012). “A fast computation of pairwise sequence alignment scores between a protein and a set of single-locus variants of another protein,” in *Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine-BCB*, Orlando, FL. doi: 10.1145/2382936.2382989
- Choi, Y., and Chan, A. P. (2015). PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* 31, 2745–2747. doi: 10.1093/bioinformatics/btv195
- Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., and Chan, A. P. (2012). Predicting the functional effect of amino acid substitutions and indels. *PLoS One* 7:e46688. doi: 10.1371/journal.pone.0046688
- Chung, M.-S., Lee, S., Min, J.-H., Huang, P., Ju, H.-W., and Kim, C. S. (2016). Regulation of *Arabidopsis thaliana* plasma membrane glucose-responsive regulator (AtPGR) expression by *A. thaliana* storekeeper-like transcription factor, AtSTKL, modulates glucose response in *Arabidopsis*. *Plant Physiol. Biochem.* 104, 155–164. doi: 10.1016/j.plaphy.2016.03.029

- Cosner, M. E., Jansen, R. K., Palmer, J. D., and Downie, S. R. (1997). The highly rearranged chloroplast genome of *Trachelium caeruleum* (Campanulaceae): multiple inversions, inverted repeat expansion and contraction, transposition, insertions/deletions, and several repeat families. *Curr. Genet.* 31, 419–429. doi: 10.1007/s002940050225
- Curaba, J., Herzog, M., and Vachon, G. (2003). *GeBP*, the first member of a new gene family in *Arabidopsis*, encodes a nuclear protein with DNA-binding activity and is regulated by *KNAT1*. *Plant J.* 33, 305–317. doi: 10.1046/j.1365-313X.2003.01622.x
- Curran, A. R., and Engelman, D. M. (2003). Sequence motifs, polar interactions and conformational changes in helical membrane proteins. *Curr. Opin. Struct. Biol.* 13, 412–417. doi: 10.1016/S0959-440X(03)00102-7
- Dai, X., Zhou, L., Zhang, W., Cai, L., Guo, H., Tian, H., et al. (2016). A single amino acid substitution in the R3 domain of GLABRA1 leads to inhibition of trichome formation in *Arabidopsis* without affecting its interaction with GLABRA3. *Plant Cell Environ.* 39, 897–907. doi: 10.1111/pce.12695
- Darling, A. E., Mau, B., and Perna, N. T. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5:e11147. doi: 10.1371/journal.pone.0011147
- Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9:772. doi: 10.1038/nmeth.2109
- David, A., Razali, R., Wass, M. N., and Sternberg, M. J. E. (2012). Protein-protein interaction sites are hot spots for disease-associated nonsynonymous SNPs. *Hum. Mutat.* 33, 359–363. doi: 10.1002/humu.21656
- de Jong, M., Wolters-Arts, M., García-Martínez, J. L., Mariani, C., and Vriezen, W. H. (2010). The *Solanum lycopersicum* AUXIN RESPONSE FACTOR 7 (SlARF7) mediates cross-talk between auxin and gibberellin signalling during tomato fruit set and development. *J. Exp. Bot.* 62, 617–626. doi: 10.1093/jxb/erq293
- de Jong, M., Wolters-Arts, M., Schimmel, B. C. J., Stultiens, C. L. M., de Groot, P. F. M., Powers, S. J., et al. (2015). *Solanum lycopersicum* AUXIN RESPONSE FACTOR 9 regulates cell division activity during early tomato fruit development. *J. Exp. Bot.* 66, 3405–3416. doi: 10.1093/jxb/erv152
- Domazet-Lošo, T. (2003). An evolutionary analysis of orphan genes in *Drosophila*. *Genome Res.* 13, 2213–2219. doi: 10.1101/gr.1311003
- Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113. doi: 10.1186/1471-2105-5-113
- Eisenberg, D., Weiss, R. M., and Terwilliger, T. C. (1982). The helical hydrophobic moment: a measure of the amphiphilicity of a helix. *Nature* 299, 371–374. doi: 10.1038/299371a0
- Ezer, D., Zabet, N. R., and Adryan, B. (2014). Homotypic clusters of transcription factor binding sites: a model system for understanding the physical mechanics of gene expression. *Comput. Struct. Biotechnol. J.* 10, 63–69. doi: 10.1016/j.csbj.2014.07.005
- Farré, D., Roset, R., Huerta, M., Adsua, J. E., Roselló, L., Albà, M. M., et al. (2003). Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. *Nucleic Acids Res.* 31, 3651–3653. doi: 10.1093/nar/kgk605
- Fei, Z., Tang, X., Alba, R. M., White, J. A., Ronning, C. M., Martin, G. B., et al. (2004). Comprehensive EST analysis of tomato and comparative genomics of fruit ripening. *Plant J.* 40, 47–59. doi: 10.1111/j.1365-313X.2004.02188.x
- Ferrándiz, C., Gu, Q., Martienssen, R., and Yanofsky, M. F. (2000). Redundant regulation of meristem identity and plant architecture by FRUITFULL, APETALA1 and CAULIFLOWER. *Development* 127, 725–734.
- Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151, 1531–1545.
- Fourquin, C., del Cerro, C., Victoria, F. C., Viallette-Guiraud, A., de Oliveira, A. C., and Ferrándiz, C. (2013). A change in SHATTERPROOF protein lies at the origin of a fruit morphological novelty and a new strategy for seed dispersal in *Medicago* genus. *Plant Physiol.* 162, 907–917. doi: 10.1104/pp.113.2.17570
- Fu, D., Szűcs, P., Yan, L., Helguera, M., Skinner, J. S., von Zitzewitz, J., et al. (2005). Large deletions within the first intron in VRN-1 are associated with spring growth habit in barley and wheat. *Mol. Genet. Genomics* 274, 442–443. doi: 10.1007/s00438-005-0045-0
- Gao, R., Wang, Y., Gruber, M. Y., and Hannoufa, A. (2018). miR156/SPL10 modulates lateral root development, branching and leaf morphology in *Arabidopsis* by silencing *AGAMOUS-LIKE 79*. *Front. Plant Sci.* 8:2226. doi: 10.3389/fpls.2017.02226
- Gazzani, S., Gendall, A. R., Lister, C., and Dean, C. (2003). Analysis of the molecular basis of flowering time variation in *Arabidopsis* accessions. *Plant Physiol.* 132, 1107–1114. doi: 10.1104/pp.103.021212
- Gillaspy, G., Ben-David, H., and Gruissem, W. (1993). Fruits: a developmental perspective. *Plant Cell* 5, 1439–1451. doi: 10.1105/tpc.5.10.1439
- Giorgetti, L., Siggers, T., Tiana, G., Caprara, G., Notarbartolo, S., Corona, T., et al. (2010). Noncooperative interactions between transcription factors and clustered DNA binding sites enable graded transcriptional responses to environmental inputs. *Mol. Cell* 37, 418–428. doi: 10.1016/j.molcel.2010.01.016
- 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, 644–652. doi: 10.1038/nbt.1883
- Gu, Q., Ferrándiz, C., Yanofsky, M. F., and Martienssen, R. (1998). The FRUITFULL MADS-box gene mediates cell differentiation during *Arabidopsis* fruit development. *Development* 125, 1509–1517.
- Guertin, M. J., and Lis, J. T. (2010). Chromatin landscape dictates HSF binding to target DNA elements. *PLoS Genet.* 6:e1001114. doi: 10.1371/journal.pgen.1001114
- Haberer, G., Hindemitt, T., Meyers, B. C., and Mayer, K. F. X. (2004). Transcriptional similarities, dissimilarities, and conservation of cis-elements in duplicated genes of *Arabidopsis*. *Plant Physiol.* 136, 3009–3022. doi: 10.1104/pp.104.046466
- Haberle, R. C., Fourcade, H. M., Boore, J. L., and Jansen, R. K. (2008). Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. *J. Mol. Evol.* 66, 350–361. doi: 10.1007/s00239-008-9086-4
- Hanzawa, Y., Money, T., and Bradley, D. (2005). A single amino acid converts a repressor to an activator of flowering. *Proc. Natl. Acad. Sci. U.S.A.* 102, 7748–7753. doi: 10.1073/pnas.0500932102
- Heijmans, K., Morel, P., and Vandenbussche, M. (2012). MADS-box genes and floral development: the dark side. *J. Exp. Bot.* 63, 5397–5404. doi: 10.1093/jxb/ers233
- Hichri, I., Deluc, L., Barrieu, F., Bogs, J., Mahjoub, A., Regad, F., et al. (2011). A single amino acid change within the R2 domain of the VvMYB5b transcription factor modulates affinity for protein partners and target promoters selectivity. *BMC Plant Biol.* 11:117. doi: 10.1186/1471-2229-11-117
- Hileman, L. C., Sundstrom, J. F., Litt, A., Chen, M., Shumba, T., and Irish, V. F. (2006). Molecular and phylogenetic analyses of the MADS-box gene family in tomato. *Mol. Biol. Evol.* 23, 2245–2258. doi: 10.1093/molbev/msl095
- Hoekstra, H. E., Hirschmann, R. J., Bunday, R. A., Insel, P. A., and Crossland, J. P. (2006). A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* 313, 101–104. doi: 10.1126/science.1126121
- Hulse-Kemp, A. M., Maheshwari, S., Stoffel, K., Hill, T. A., Jaffe, D., Williams, S. R., et al. (2018). Reference quality assembly of the 3.5-Gb genome of *Capsicum annuum* from a single linked-read library. *Hortic. Res.* 5:4. doi: 10.1038/s41438-017-0011-0
- Hwang, Y.-S., Cicci, P., Parsons, R. L., Moose, S. P., Schmidt, R. J., and Huang, N. (2004). The maize O2 and PBF proteins act additively to promote transcription from storage protein gene promoters in rice endosperm cells. *Plant Cell Physiol.* 45, 1509–1518. doi: 10.1093/pcp/pch173
- Immink, R. G. H., Kaufmann, K., and Angenent, G. C. (2010). The “ABC” of MADS domain protein behaviour and interactions. *Semin. Cell Dev. Biol.* 21, 87–93. doi: 10.1016/j.semcdb.2009.10.004
- Ingram, V. M. (1957). Gene mutations in human haemoglobin: the chemical difference between normal and sickle cell haemoglobin. *Nature* 180, 326–328. doi: 10.1038/180326a0
- Jacobsen, M. W., da Fonseca, R. R., Bernatchez, L., and Hansen, M. M. (2016). Comparative analysis of complete mitochondrial genomes suggests that relaxed purifying selection is driving high nonsynonymous evolutionary rate of the NADH2 gene in whitefish (*Coregonus* spp.). *Mol. Phylogenet. Evol.* 95, 161–170. doi: 10.1016/j.ympev.2015.11.008

- Jeffares, D. C., Pain, A., Berry, A., Cox, A. V., Stalker, J., Ingle, C. E., et al. (2006). Genome variation and evolution of the malaria parasite *Plasmodium falciparum*. *Nat. Genet.* 39, 120–125. doi: 10.1038/ng1931
- Jun, J. H., Liu, C., Xiao, X., and Dixon, R. A. (2015). The transcriptional repressor MYB2 regulates both spatial and temporal patterns of proanthocyanidin and anthocyanin pigmentation in *Medicago truncatula*. *Plant Cell* 27, 2860–2879. doi: 10.1105/tpc.15.00476
- Kaufmann, K., Melzer, R., and Theissen, G. (2005). MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. *Gene* 347, 183–198. doi: 10.1016/j.gene.2004.12.014
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858. doi: 10.1038/nprot.2015.053
- Kim, S., Koh, J., Yoo, M.-J., Kong, H., Hu, Y., Ma, H., et al. (2005). Expression of floral MADS-box genes in basal angiosperms: implications for the evolution of floral regulators. *Plant J.* 43, 724–744. doi: 10.1111/j.1365-313X.2005.02487.x
- Kmita, M., Kondo, T., and Duboule, D. (2000). Targeted inversion of a polar silencer within the HoxD complex re-allocates domains of enhancer sharing. *Nat. Genet.* 26, 451–454. doi: 10.1038/82593
- Knapp, S. (2002). Tobacco to tomatoes: a phylogenetic perspective on fruit diversity in the Solanaceae. *J. Exp. Bot.* 53, 2001–2022. doi: 10.1093/jxb/erf068
- Krueger, F. (2017). *Trim Galore!*. Available at: http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
- Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 30, 3276–3278. doi: 10.1093/bioinformatics/btu531
- Lebrecht, D., Foehr, M., Smith, E., Lopes, F. J. P., Vanario-Alonso, C. E., Reinitz, J., et al. (2005). Bicoid cooperative DNA binding is critical for embryonic patterning in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* 102, 13176–13181. doi: 10.1073/pnas.0506462102
- Lee, S., Chung, E.-J., Joong, Y.-H., and Choi, D. (2010). Non-climacteric fruit ripening in pepper: increased transcription of EIL-like genes normally regulated by ethylene. *Funct. Integr. Genomics* 10, 135–146. doi: 10.1007/s10142-009-0136-9
- Lercher, M. J., Blumenthal, T., and Hurst, L. D. (2003). Coexpression of neighboring genes in *Caenorhabditis elegans* is mostly due to operons and duplicate genes. *Genome Res.* 13, 238–243. doi: 10.1101/gr.553803
- Levo, M., Zalcvar, E., Sharon, E., Dantas Machado, A. C., Kalma, Y., Lotam-Pompan, M., et al. (2015). Unraveling determinants of transcription factor binding outside the core binding site. *Genome Res.* 25, 1018–1029. doi: 10.1101/gr.185033.114
- Li, M., Petukh, M., Alexov, E., and Panchenko, A. R. (2014). Predicting the impact of missense mutations on protein–protein binding affinity. *J. Chem. Theory Comput.* 10, 1770–1780. doi: 10.1021/ct401022c
- Liljegren, S. J., Ditta, G. S., Eshed, Y., Savidge, B., Bowman, J. L., and Yanofsky, M. F. (2000). SHATTERPROOF MADS-box genes control seed dispersal in *Arabidopsis*. *Nature* 404, 766–770. doi: 10.1038/35008089
- Liljegren, S. J., Roeder, A. H. K., Kempin, S. A., Gremski, K., Østergaard, L., Guimil, S., et al. (2004). Control of fruit patterning in *Arabidopsis* by INDEHISCENT. *Cell* 116, 843–853. doi: 10.1016/S0092-8674(04)00217-X
- Litt, A., and Irish, V. F. (2003). Duplication and diversification in the APETALA1/FRUITFULL floral homeotic gene lineage: implications for the evolution of floral development. *Genetics* 165, 821–833.
- Liu, M., Watson, L. T., and Zhang, L. (2014). Quantitative prediction of the effect of genetic variation using hidden Markov models. *BMC Bioinformatics* 15:5. doi: 10.1186/1471-2105-15-5
- Liu, S., Zhang, Y., Feng, Q., Qin, L., Pan, C., Lamin-Samu, A. T., et al. (2018). Tomato AUXIN RESPONSE FACTOR 5 regulates fruit set and development via the mediation of auxin and gibberellin signaling. *Sci. Rep.* 8:2971. doi: 10.1038/s41598-018-21315-y
- Liu, X., Lee, C.-K., Granek, J. A., Clarke, N. D., and Lieb, J. D. (2006). Whole-genome comparison of Leu3 binding in vitro and in vivo reveals the importance of nucleosome occupancy in target site selection. *Genome Res.* 16, 1517–1528. doi: 10.1101/gr.5655606
- Lupiáñez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., et al. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* 161, 1012–1025. doi: 10.1016/j.cell.2015.04.004
- Lupski, J. R. (1998). Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet.* 14, 417–422. doi: 10.1016/S0168-9525(98)01555-8
- Ma, B., Elkayam, T., Wolfson, H., and Nussinov, R. (2003). Protein-protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces. *Proc. Natl. Acad. Sci. U.S.A.* 100, 5772–5777. doi: 10.1073/pnas.1030237100
- Martinez-Castilla, L. P., and Alvarez-Buylla, E. R. (2003). Adaptive evolution in the *Arabidopsis* MADS-box gene family inferred from its complete resolved phylogeny. *Proc. Natl. Acad. Sci. U.S.A.* 100, 13407–13412. doi: 10.1073/pnas.1835864100
- Massa, A. N., Childs, K. L., Lin, H., Bryan, G. J., Giuliano, G., and Buell, C. R. (2011). The transcriptome of the reference potato genome *Solanum tuberosum* Group Phureja clone DM1-3 516R44. *PLoS One* 6:e26801. doi: 10.1371/journal.pone.0026801
- 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* 3:17. doi: 10.1186/2047-217X-3-17
- Melzer, S., Lens, F., Gennen, J., Vanneste, S., Rohde, A., and Beeckman, T. (2008). Flowering-time genes modulate meristem determinacy and growth form in *Arabidopsis thaliana*. *Nat. Genet.* 40, 1489–1492. doi: 10.1038/ng.253
- Messegueur, X., Escudero, R., Farré, D., Núñez, O., Martínez, J., and Albà, M. M. (2002). PROMO: detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics* 18, 333–334. doi: 10.1093/bioinformatics/18.2.333
- Michaels, S. D., He, Y., Scortecci, K. C., and Amasino, R. M. (2003). Attenuation of FLOWERING LOCUS C activity as a mechanism for the evolution of summer-annual flowering behavior in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 100, 10102–10107. doi: 10.1073/pnas.1531467100
- Müller, B. M., Saedler, H., and Zachgo, S. (2001). The MADS-box gene *DEFH28* from *Antirrhinum* is involved in the regulation of floral meristem identity and fruit development. *Plant J.* 28, 169–179. doi: 10.1046/j.1365-313X.2001.01139.x
- Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., and Kosakovsky Pond, S. L. (2012). Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* 8:e1002764. doi: 10.1371/journal.pgen.1002764
- Nakasugi, K., Crowhurst, R., Bally, J., and Waterhouse, P. (2014). Combining transcriptome assemblies from multiple de novo assemblers in the allo-tetraploid plant *Nicotiana benthamiana*. *PLoS One* 9:e91776. doi: 10.1371/journal.pone.0091776
- NCBI Resource Coordinators (2017). Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 45, D12–D17. doi: 10.1093/nar/gkw1071
- Nei, M., Gu, X., and Sitnikova, T. (1997). Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc. Natl. Acad. Sci. U.S.A.* 94, 7799–7806. doi: 10.1073/pnas.94.15.7799
- Nielsen, R., Bustamante, C., Clark, A. G., Glanowski, S., Sackton, T. B., Hubisz, M. J., et al. (2005). A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* 3:e170. doi: 10.1371/journal.pbio.0030170
- Nietzsche, M., Guerra, T., Alseekh, S., Wiermer, M., Sonnewald, S., Fernie, A. R., et al. (2017). STROKEKEEPER RELATED1/G-element binding protein (STKR1) interacts with protein kinase SnRK1. *Plant Physiol.* 176, 1773–1792. doi: 10.1104/pp.17.01461
- Olmstead, R. G., Bohs, L., Migid, H. A., Santiago-Valentin, E., Garcia, V. F., and Collier, S. M. (2008). A molecular phylogeny of the Solanaceae. *Taxon* 57, 1159–1181.
- Ortiz-Ramírez, C. I., Plata-Arboleda, S., and Pabón-Mora, N. (2018). Evolution of genes associated with gynoecium patterning and fruit development in Solanaceae. *Ann. Bot.* 121, 1211–1230. doi: 10.1093/aob/mcy007
- Ota, T., and Nei, M. (1994). Divergent evolution and evolution by the birth-and-death process in the immunoglobulin VH gene family. *Mol. Biol. Evol.* 11, 469–482.

- Pabón-Mora, N., Ambrose, B. A., and Litt, A. (2012). Poppy APETALA1/FRUITFULL orthologs control flowering time, branching, perianth identity, and fruit development. *Plant Physiol.* 158, 1685–1704. doi: 10.1104/pp.111.192104
- Pabón-Mora, N., Sharma, B., Holappa, L. D., Kramer, E. M., and Litt, A. (2013). The Aquilegia FRUITFULL-like genes play key roles in leaf morphogenesis and inflorescence development. *Plant J.* 74, 197–212. doi: 10.1111/tpj.12113
- Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20, 289–290. doi: 10.1093/bioinformatics/btg412
- Payne, J. L., and Wagner, A. (2015). Mechanisms of mutational robustness in transcriptional regulation. *Front. Genet.* 6:322. doi: 10.3389/fgene.2015.00322
- Pegueroles, C., Laurie, S., and Albà, M. M. (2013). Accelerated evolution after gene duplication: a time-dependent process affecting just one copy. *Mol. Biol. Evol.* 30, 1830–1842. doi: 10.1093/molbev/mst083
- Pellegrini, L., Tan, S., and Richmond, T. J. (1995). Structure of serum response factor core bound to DNA. *Nature* 376, 490–498. doi: 10.1038/376490a0
- Piontkivska, H., Rooney, A. P., and Nei, M. (2002). Purifying selection and birth-and-death evolution in the histone H4 gene family. *Mol. Biol. Evol.* 19, 689–697. doi: 10.1093/oxfordjournals.molbev.a004127
- Potato Genome Sequencing Consortium, Xu, X., Pan, S., Cheng, S., Zhang, B., Mu, D., et al. (2011). Genome sequence and analysis of the tuber crop potato. *Nature* 475, 189–195. doi: 10.1038/nature10158
- Preston, J. C., and Kellogg, E. A. (2007). Conservation and divergence of APETALA1/FRUITFULL-like gene function in grasses: evidence from gene expression analyses. *Plant J.* 52, 69–81. doi: 10.1111/j.1365-313X.2007.03209.x
- Preston, J. C., and Kellogg, E. A. (2008). Discrete developmental roles for temperate cereal grass VERNALIZATION1/FRUITFULL-like genes in flowering competency and the transition to flowering. *Plant Physiol.* 146, 265–276. doi: 10.1104/pp.107.109561
- Prince, V. E., and Pickett, F. B. (2002). Splitting pairs: the diverging fates of duplicated genes. *Nat. Rev. Genet.* 3, 827–837. doi: 10.1038/nrg928
- Puig, M., Casillas, S., Villatoro, S., and Cáceres, M. (2015). Human inversions and their functional consequences. *Brief. Funct. Genomics* 14, 369–379. doi: 10.1093/bfpg/elv020
- Purugganan, M. D., Rounsley, S. D., Schmidt, R. J., and Yanofsky, M. F. (1995). Molecular evolution of flower development: diversification of the plant MADS-box regulatory gene family. *Genetics* 140, 345–356.
- Python Language Reference (2010). *Python Software Foundation*. Available at: <http://www.python.org>
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Rajani, S., and Sundaresan, V. (2001). The *Arabidopsis* myc/bHLH gene ALCATRAZ enables cell separation in fruit dehiscence. *Curr. Biol.* 11, 1914–1922. doi: 10.1016/S0960-9822(01)00593-0
- Robinson, D. F., and Foulds, L. R. (1981). Comparison of phylogenetic trees. *Math. Biosci.* 53, 131–147. doi: 10.1016/0025-5564(81)90043-2
- Sakuma, S., Lundqvist, U., Kakei, Y., Thirulogachandar, V., Suzuki, T., Hori, K., et al. (2017). Extreme suppression of lateral fleret development by a single amino acid change in the VRS1 transcription factor. *Plant Physiol.* 175, 1720–1731. doi: 10.1104/pp.17.01149
- Särkinen, T., Bohs, L., Olmstead, R. G., and Knapp, S. (2013). A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. *BMC Evol. Biol.* 13:214. doi: 10.1186/1471-2148-13-214
- Schauer, S. E., Schlüter, P. M., Baskar, R., Gheyselinck, J., Bolaños, A., Curtis, M. D., et al. (2009). Intronic regulatory elements determine the divergent expression patterns of AGAMOUS-LIKE6 subfamily members in Arabidopsis. *Plant J.* 59, 987–1000. doi: 10.1111/j.1365-313X.2009.03928.x
- Schliep, K. P. (2010). phangorn: phylogenetic analysis in R. *Bioinformatics* 27, 592–593. doi: 10.1093/bioinformatics/btq706
- Schlueter, J. A., Dixon, P., Granger, C., Grant, D., Clark, L., Doyle, J. J., et al. (2004). Mining EST databases to resolve evolutionary events in major crop species. *Genome* 47, 868–876. doi: 10.1139/g04-047
- Schröfelbauer, B., Chen, D., and Landau, N. R. (2004). A single amino acid of APOBEC3G controls its species-specific interaction with virion infectivity factor (Vif). *Proc. Natl. Acad. Sci. U.S.A.* 101, 3927–3932. doi: 10.1073/pnas.0307132101
- Scorza, L. C. T., Hernandez-Lopes, J., Melo-de-Pinna, G. F. A., and Dornelas, M. C. (2017). Expression patterns of *Passiflora edulis* APETALA1/FRUITFULL homologues shed light onto tendril and corona identities. *Evodevo* 8:3. doi: 10.1186/s13227-017-0066-x
- Shan, H., Zhang, N., Liu, C., Xu, G., Zhang, J., Chen, Z., et al. (2007). Patterns of gene duplication and functional diversification during the evolution of the AP1/SQUA subfamily of plant MADS-box genes. *Mol. Phylogenet. Evol.* 44, 26–41. doi: 10.1016/j.ympev.2007.02.016
- Shchennikova, A. V., Shulga, O. A., Immink, R., Skryabin, K. G., and Angenent, G. C. (2004). Identification and characterization of four chrysanthemum MADS-box genes, belonging to the APETALA1/FRUITFULL and SEPALLATA3 subfamilies. *Plant Physiol.* 134, 1632–1641. doi: 10.1104/pp.103.036665
- Sheinerman, F. B., Norel, R., and Honig, B. (2000). Electrostatic aspects of protein-protein interactions. *Curr. Opin. Struct. Biol.* 10, 153–159. doi: 10.1016/S0959-440X(00)00065-8
- Shima, Y., Fujisawa, M., Kitagawa, M., Nakano, T., Kimbara, J., Nakamura, N., et al. (2014). Tomato FRUITFULL homologs regulate fruit ripening via ethylene biosynthesis. *Biosci. Biotechnol. Biochem.* 78, 231–237. doi: 10.1080/09168451.2014.878221
- Shima, Y., Kitagawa, M., Fujisawa, M., Nakano, T., Kato, H., Kimbara, J., et al. (2013). Tomato FRUITFULL homologues act in fruit ripening via forming MADS-box transcription factor complexes with RIN. *Plant Mol. Biol.* 82, 427–438. doi: 10.1007/s11103-013-0071-y
- Sierro, N., Battey, J. N. D., Ouadi, S., Bovet, L., Goepfert, S., Bakaher, N., et al. (2013). Reference genomes and transcriptomes of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*. *Genome Biol.* 14:R60. doi: 10.1186/gb-2013-14-6-r60
- Slugina, M. A., Shchennikova, A. V., Pishnaya, O. N., and Kochieva, E. Z. (2018). Assessment of the fruit-ripening-related *FUL2* gene diversity in morphophysiologically contrasted cultivated and wild tomato species. *Mol. Breed.* 38:82. doi: 10.1007/s11032-018-0842-x
- Smykal, P., Gennen, J., De Bodt, S., Ranganath, V., and Melzer, S. (2007). Flowering of strict photoperiodic *Nicotiana* varieties in non-inductive conditions by transgenic approaches. *Plant Mol. Biol.* 65, 233–242. doi: 10.1007/s11103-007-9211-6
- Song, C., Guo, J., Sun, W., and Wang, Y. (2012). Whole genome duplication of intra- and inter-chromosomes in the tomato genome. *J. Genet. Genomics* 39, 361–368. doi: 10.1016/j.jgg.2012.06.002
- Spence, J., Vercher, Y., Gates, P., and Harris, N. (1996). “Pod shatter” in *Arabidopsis thaliana*, *Brassica napus* and *B. juncea*. *J. Microsc.* 181, 195–203. doi: 10.1046/j.1365-2818.1996.111391.x
- Stefanović, S., Austin, D. F., and Olmstead, R. G. (2003). Classification of Convolvulaceae: a phylogenetic approach. *Syst. Bot.* 28, 791–806.
- Sukumaran, J., and Holder, M. T. (2010). DendroPy: a Python library for phylogenetic computing. *Bioinformatics* 26, 1569–1571. doi: 10.1093/bioinformatics/btq228
- Takumi, S., Koyama, K., Fujiwara, K., and Kobayashi, F. (2011). Identification of a large deletion in the first intron of the *Vrn-D1* locus, associated with loss of vernalization requirement in wild wheat progenitor *Aegilops tauschii* Coss. *Genes Genet. Syst.* 86, 183–195. doi: 10.1266/ggs.86.183
- Tanimoto, K., Liu, Q., Bungert, J., and Engel, J. D. (1999). Effects of altered gene order or orientation of the locus control region on human β -globin gene expression in mice. *Nature* 398, 344–348. doi: 10.1038/18698
- Tanksley, S. D. (2004). The genetic, developmental, and molecular bases of fruit size and shape variation in tomato. *Plant Cell* 16(Suppl.), S181–S189. doi: 10.1105/tpc.018119
- Teng, S., Madej, T., Panchenko, A., and Alexov, E. (2009). Modeling effects of human single nucleotide polymorphisms on protein-protein interactions. *Biophys. J.* 96, 2178–2188. doi: 10.1016/j.bpj.2008.12.3904
- The Tomato Genome Consortium (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485, 635–641. doi: 10.1038/nature11119
- Tieman, D. M., Ciardi, J. A., Taylor, M. G., and Klee, H. J. (2001). Members of the tomato LeEIL (EIN3-like) gene family are functionally redundant and

- regulate ethylene responses throughout plant development. *Plant J.* 26, 47–58. doi: 10.1046/j.1365-313x.2001.01006.x
- Torgerson, D. G., Kulathinal, R. J., and Singh, R. S. (2002). Mammalian sperm proteins are rapidly evolving: evidence of positive selection in functionally diverse genes. *Mol. Biol. Evol.* 19, 1973–1980. doi: 10.1093/oxfordjournals.molbev.a004021
- Vanneste, K., Maere, S., and Van de Peer, Y. (2014). Tangled up in two: a burst of genome duplications at the end of the Cretaceous and the consequences for plant evolution. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20130353. doi: 10.1098/rstb.2013.0353
- Vision, T. J., Brown, D. G., and Tanksley, S. D. (2000). The origins of genomic duplications in *Arabidopsis*. *Science* 290, 2114–2117. doi: 10.1126/science.290.5499.2114
- Vogel, M. J., Pagie, L., Talhout, W., Nieuwland, M., Kerkhoven, R. M., and van Steensel, B. (2009). High-resolution mapping of heterochromatin redistribution in a *Drosophila* position-effect variegation model. *Epigenetics Chromatin* 2:1. doi: 10.1186/1756-8935-2-1
- Wang, S., Lu, G., Hou, Z., Luo, Z., Wang, T., Li, H., et al. (2014). Members of the tomato *FRUITFULL* MADS-box family regulate style abscission and fruit ripening. *J. Exp. Bot.* 65, 3005–3014. doi: 10.1093/jxb/eru137
- White, M. A., Myers, C. A., Corbo, J. C., and Cohen, B. A. (2013). Massively parallel in vivo enhancer assay reveals that highly local features determine the cis-regulatory function of ChIP-seq peaks. *Proc. Natl. Acad. Sci. U.S.A.* 110, 11952–11957. doi: 10.1073/pnas.1307449110
- Xu, Q., Yin, X.-R., Zeng, J.-K., Ge, H., Song, M., Xu, C.-J., et al. (2014). Activator- and repressor-type MYB transcription factors are involved in chilling injury induced flesh lignification in loquat via their interactions with the phenylpropanoid pathway. *J. Exp. Bot.* 65, 4349–4359. doi: 10.1093/jxb/eru208
- Yang, Y., Fanning, L., and Jack, T. (2003a). The K domain mediates heterodimerization of the *Arabidopsis* floral organ identity proteins, *APETALA3* and *PISTILLATA*. *Plant J.* 33, 47–59.
- Yang, Y., Xiang, H., and Jack, T. (2003b). *pistillata-5*, an *Arabidopsis* B class mutant with strong defects in petal but not in stamen development. *Plant J.* 33, 177–188.
- Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* 13, 555–556. doi: 10.1093/bioinformatics/13.5.555
- Yang, Z., and Bielawski, J. P. (2000). Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* 15, 496–503. doi: 10.1016/S0169-5347(00)01994-7
- 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
- Yu, J., Wang, L., Guo, H., Liao, B., King, G., and Zhang, X. (2017). Genome evolutionary dynamics followed by diversifying selection explains the complexity of the *Sesamum indicum* genome. *BMC Genomics* 18:257. doi: 10.1186/s12864-017-3599-4
- Zhao, H., Wang, X., Zhu, D., Cui, S., Li, X., Cao, Y., et al. (2012). A single amino acid substitution in IIIf subfamily of basic helix-loop-helix transcription factor AtMYC1 leads to trichome and root hair patterning defects by abolishing its interaction with partner proteins in *Arabidopsis*. *J. Biol. Chem.* 287, 14109–14121. doi: 10.1074/jbc.M111.280735
- Zhou, P., Hou, S., Bai, Z., Li, Z., Wang, H., Chen, Z., et al. (2018). Disrupting the intramolecular interaction between proto-oncogene c-Src SH3 domain and its self-binding peptide PPII with rationally designed peptide ligands. *Artif. Cells Nanomed. Biotechnol.* 46, 1122–1131. doi: 10.1080/21691401.2017.1360327
- Zourelidou, M., de Torres-Zabala, M., Smith, C., and Bevan, M. W. (2002). Storekeeper defines a new class of plant-specific DNA-binding proteins and is a putative regulator of patatin expression. *Plant J.* 30, 489–497. doi: 10.1046/j.1365-313X.2002.01302.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 Maheepala, Emerling, Rajewski, Macon, Strahl, Pabón-Mora and Litt. 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