

# INTEGRATION OF OMICS DATA TO UNDERSTAND PLANT METABOLISM

EDITED BY: Carlos Alberto Labate, Diego Mauricio Riaño-Pachón,  
Glória Catarina Pinto, Paulo Mazzafera and Luis Valledor  
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 978-2-88945-910-0

DOI 10.3389/978-2-88945-910-0

## 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](mailto:researchtopics@frontiersin.org)



# INTEGRATION OF OMICS DATA TO UNDERSTAND PLANT METABOLISM

Topic Editors:

**Carlos Alberto Labate**, University of São Paulo, Brazil

**Diego Mauricio Riaño-Pachón**, University of São Paulo, Brazil

**Glória Catarina Pinto**, University of Aveiro, Portugal

**Paulo Mazzafera**, Campinas State University, Brazil

**Luis Valledor**, Universidad de Oviedo, Spain

**Citation:** Labate, C. A., Riaño-Pachón, D. M., Pinto, G. C., Mazzafera, P., Valledor, L., eds. (2019). Integration of OMICS Data to Understand Plant Metabolism.

Lausanne: Frontiers Media. doi: 10.3389/978-2-88945-910-0



# Table of Contents

- 04 *Eco-Metabolomics and Metabolic Modeling: Making the Leap From Model Systems in the Lab to Native Populations in the Field***  
Matthias Nagler, Thomas Nägele, Christian Gilli, Lena Fragner, Arthur Korte, Alexander Platzter, Ashley Farlow, Magnus Nordborg and Wolfram Weckwerth
- 17 *A Multi-Omics Analysis Pipeline for the Metabolic Pathway Reconstruction in the Orphan Species Quercus ilex***  
Cristina López-Hidalgo, Victor M. Guerrero-Sánchez, Isabel Gómez-Gálvez, Rosa Sánchez-Lucas, María A. Castillejo-Sánchez, Ana M. Maldonado-Alconada, Luis Valledor and Jesus V. Jorrín-Novó
- 33 *Metabolite Profiles of Sugarcane Culm Reveal the Relationship Among Metabolism and Axillary Bud Outgrowth in Genetically Related Sugarcane Commercial Cultivars***  
Danilo A. Ferreira, Marina C. M. Martins, Adriana Cheavegatti-Gianotto, Monalisa S. Carneiro, Rodrigo R. Amadeu, Juliana A. Aricetti, Lucia D. Wolf, Hermann P. Hoffmann, Luis G. F. de Abreu and Camila Caldana
- 47 *Combined Drought and Heat Activates Protective Responses in Eucalyptus globulus That are not Activated When Subjected to Drought or Heat Stress Alone***  
Barbara Correia, Robert D. Hancock, Joana Amaral, Aurelio Gomez-Cadenas, Luis Valledor and Glória Pinto
- 61 *Molecular Profiling of Pierce's Disease Outlines the Response Circuitry of Vitis vinifera to Xylella fastidiosa Infection***  
Paulo A. Zaini, Rafael Nascimento, Hossein Gouran, Dario Cantu, Sandeep Chakraborty, My Phu, Luiz R. Goulart and Abhaya M. Dandekar
- 77 *Metabolome Integrated Analysis of High-Temperature Response in Pinus radiata***  
Mónica Escandón, Mónica Meijón, Luis Valledor, Jesús Pascual, Gloria Pinto and María Jesús Cañal
- 92 *Cyclotide Evolution: Insights From the Analyses of Their Precursor Sequences, Structures and Distribution in Violets (Viola)***  
Sungkyu Park, Ki-Oug Yoo, Thomas Marcussen, Anders Backlund, Erik Jacobsson, K. Johan Rosengren, Inseok Doo and Ulf Göransson
- 111 *The Challenge to Translate OMICS Data to Whole Plant Physiology: The Context Matters***  
Marcelo N. do Amaral and Gustavo M. Souza
- 115 *Inference of Transcription Regulatory Network in Low Phytic Acid Soybean Seeds***  
Neelam Redekar, Guillaume Pilot, Victor Raboy, Song Li and M. A. Saghai Maroof
- 129 *An Integrated "Multi-Omics" Comparison of Embryo and Endosperm Tissue-Specific Features and Their Impact on Rice Seed Quality***  
Marc Galland, Dongli He, Imen Lounifi, Erwann Arc, Gilles Clément, Sandrine Balzergue, Stéphanie Huguet, Gwendal Cueff, Béatrice Godin, Boris Collet, Fabienne Granier, Halima Morin, Joseph Tran, Benoit Valot and Loïc Rajjou





# Eco-Metabolomics and Metabolic Modeling: Making the Leap From Model Systems in the Lab to Native Populations in the Field

Matthias Nagler<sup>1†</sup>, Thomas Nägele<sup>1,2†</sup>, Christian Gilli<sup>1</sup>, Lena Fragner<sup>1,3</sup>, Arthur Korte<sup>4</sup>, Alexander Platzer<sup>5</sup>, Ashley Farlow<sup>5</sup>, Magnus Nordborg<sup>5</sup> and Wolfram Weckwerth<sup>1,3\*</sup>

<sup>1</sup> Department of Ecogenomics and Systems Biology, University of Vienna, Vienna, Austria, <sup>2</sup> LMU Munich, Plant Evolutionary Cell Biology, Munich, Germany, <sup>3</sup> Vienna Metabolomics Center (VIME), University of Vienna, Vienna, Austria, <sup>4</sup> Center for Computational and Theoretical Biology, University of Würzburg, Würzburg, Germany, <sup>5</sup> Gregor Mendel Institute of Molecular Plant Biology, Austrian Academy of Sciences, Vienna, Austria

## OPEN ACCESS

### Edited by:

Luis Valledor,  
Universidad de Oviedo, Spain

### Reviewed by:

Jesús Pascual Vázquez,  
University of Turku, Finland  
Federico Valverde,  
Consejo Superior de Investigaciones  
Científicas (CSIC), Spain

### \*Correspondence:

Wolfram Weckwerth  
wolfram.weckwerth@univie.ac.at

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Plant Systems and Synthetic Biology,  
a section of the journal  
Frontiers in Plant Science

**Received:** 30 March 2018

**Accepted:** 04 October 2018

**Published:** 06 November 2018

### Citation:

Nagler M, Nägele T, Gilli C,  
Fragner L, Korte A, Platzer A,  
Farlow A, Nordborg M and  
Weckwerth W (2018)  
Eco-Metabolomics and Metabolic  
Modeling: Making the Leap From  
Model Systems in the Lab to Native  
Populations in the Field.  
Front. Plant Sci. 9:1556.  
doi: 10.3389/fpls.2018.01556

Experimental high-throughput analysis of molecular networks is a central approach to characterize the adaptation of plant metabolism to the environment. However, recent studies have demonstrated that it is hardly possible to predict *in situ* metabolic phenotypes from experiments under controlled conditions, such as growth chambers or greenhouses. This is particularly due to the high molecular variance of *in situ* samples induced by environmental fluctuations. An approach of functional metabolome interpretation of field samples would be desirable in order to be able to identify and trace back the impact of environmental changes on plant metabolism. To test the applicability of metabolomics studies for a characterization of plant populations in the field, we have identified and analyzed *in situ* samples of nearby grown natural populations of *Arabidopsis thaliana* in Austria. *A. thaliana* is the primary molecular biological model system in plant biology with one of the best functionally annotated genomes representing a reference system for all other plant genome projects. The genomes of these novel natural populations were sequenced and phylogenetically compared to a comprehensive genome database of *A. thaliana* ecotypes. Experimental results on primary and secondary metabolite profiling and genotypic variation were functionally integrated by a data mining strategy, which combines statistical output of metabolomics data with genome-derived biochemical pathway reconstruction and metabolic modeling. Correlations of biochemical model predictions and population-specific genetic variation indicated varying strategies of metabolic regulation on a population level which enabled the direct comparison, differentiation, and prediction of metabolic adaptation of the same species to different habitats. These differences were most pronounced at organic and amino acid metabolism as well as at the interface of primary and secondary metabolism and allowed for the direct classification of population-specific metabolic phenotypes within geographically contiguous sampling sites.

**Keywords:** eco-metabolomics, *in situ* analysis, metabolomics, metabolic modeling, SNP, natural variation, Jacobian matrix, green systems biology



## INTRODUCTION

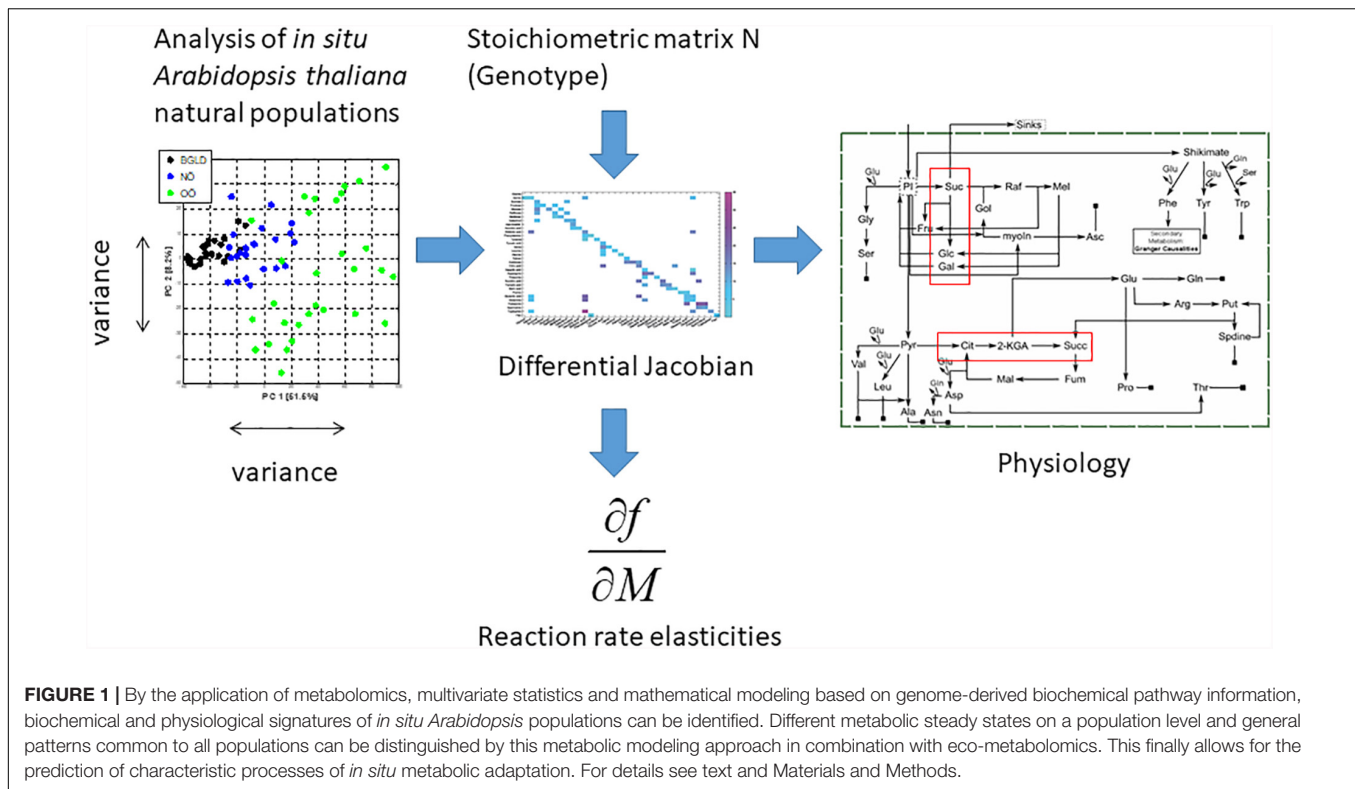
Natural variation, as first described by Darwin (1859), is the ultimate point of attack for natural selection and still the only known process that is able to produce adaptive evolutionary change. *Arabidopsis thaliana* has become a powerful model organism for studying many aspects of plant biology and adaptation to the environment (Somerville and Koornneef, 2002; Hancock et al., 2011). After the publication of a first complete reference genome sequence (Arabidopsis, 2000), it was discovered that it is inappropriate to think about ‘the’ genome of a species (Weigel and Mott, 2009). In fact, all species are exposed to specific environmental clines differently affecting individual plants’ phenotypic performance (Turesson, 1922; Ellenberg, 1953; Hoffmann, 2002; Weckwerth, 2003, 2011a; Lasky et al., 2012; Weigel, 2012). Therefore, they comprise different populations colonizing different habitats. These habitats may impose differing directions of natural selection upon the coenospecies, and thus, together with genetic drift, lead to diverging allele frequencies and to an inhomogeneous genetic structure. This inhomogeneity is called natural genetic variation and potentially provides insights in genome evolution, population structure, and selective mechanisms (Mitchell-Olds and Schmitt, 2006). However, the genetic side represents only one level in the complex molecular architecture, which builds up the basis for physiological and morphological responses of plants to environmental stimuli (Pigliucci, 2010). The experimental analysis and interpretation of these molecular architectures is nonintuitive, particularly because of the highly complex organization of plant molecular networks. Numerous studies have shown that a multitude of genes, proteins, metabolites, and underlying regulatory processes are involved in plant–environment interactions (Koornneef et al., 2004; Wienkoop et al., 2008; Keurentjes, 2009; Chan et al., 2010; Macel et al., 2010; Lasky et al., 2012). However, interpreting these findings in the context of environmental conditions and, particularly, in an ecological context is highly challenging. This is particularly due to a missing stringent definition of the genotype–phenotype relationship, which can hardly be expected to be derivable from a single methodology but rather from a comprehensive platform of experimental and theoretical strategies (Weckwerth, 2003, 2011a; Diz et al., 2012). Recording environmentally induced fluctuations in a metabolic homeostasis has been shown to be a promising approach to unravel complex patterns of metabolic regulation and adaptation. For example, the metabolism of floral anthocyanins, which is a central group of secondary metabolites, was found to represent a suitable metabolic system to characterize the process of environmental regulation (Lu et al., 2009). The authors suggested that environmental regulation of the anthocyanin pathway is mainly affected by daily average temperature and UV light intensity modulating anthocyanin transcript levels at floral developmental stages. In another study, a metabolomics approach has been applied to elucidate *in situ* allelopathic relationships of individual species to phytosociological gradients (Scherling et al., 2010). We demonstrated that *in situ* metabolic signatures of five different plant species correlated with a biodiversity gradient. More

general, metabolomics approaches can be expected to provide detailed information about metabolic processes in context of genomic signatures (Chae et al., 2014). Particularly in model systems with functionally annotated genomes this makes it the method of choice to unravel and interpret molecular ecological properties.

For the genetic and molecular biological model plant *A. thaliana*, one of the best functionally annotated genomes (Baerenfaller et al., 2012; Lavagi et al., 2012) and a comprehensive catalog of genome information is available<sup>1</sup>. Recently, an *in vitro* study of the physiological homeostasis of 92 *A. thaliana* accessions in multiple growth settings has demonstrated the devastating impact of varying environmental conditions on the correlation of *in vitro* metabolism to geographic origin (Kleessen et al., 2012). Yet, as microhabitats may vary significantly on relatively small spatial scales and are not necessarily corresponding to geographic distance, the investigation of the molecular performance of plants *in situ* seems inevitable to get a realistic picture of plant–environment interactions and their ecophysiological consequences. A well-known example indicating the need of such *in situ* studies is Ellenberg’s Hohenheimer groundwater table experiment (Ellenberg, 1953; Hector et al., 2012). Here, it was shown that the phenotypic performance of plants *in vitro* significantly differ from their *in situ* physiological homeostasis, as important microhabitat parameters may not be included in the *in vitro* growth setting (Shulaev et al., 2008). Both plant communities and plant populations seem to be an appropriate target for the development and tuning of *in situ* methodologies due to their sessile nature and the availability of a large set of *in vitro* reference data for some species. This enables the intersection of individual molecular with environmental data, and even ecosystem properties can be accounted for via geographic information systems. Genotyping approaches in *A. thaliana* have already been established (Atwell et al., 2010; Platt et al., 2010; Todesco et al., 2010; Hancock et al., 2011; Horton et al., 2012; Long et al., 2013) and are easily transferable to *in situ* samples (Hunter et al., 2013). Metabolomics and proteomics technologies provide the means for generating upstream molecular phenotypes (Morgenthal et al., 2005; Hoehenwarter et al., 2008; Wienkoop et al., 2008; Scherling et al., 2010; Weckwerth, 2011a; Doerfler et al., 2013). Thus, these techniques are suitable for experimental high-throughput analysis at the molecular level, representing the basis for strategies of multivariate statistics and mathematical modeling to identify biochemical perturbation sites and gain predictive power (Nägele and Weckwerth, 2013; Nägele, 2014). In this context, particularly metabolomic analysis has proven to be a suitable approach for the comprehensive and representative investigation of complex metabolic networks with respect to the underlying phenotypic diversity (Weckwerth et al., 2004a; Keurentjes, 2009; Scherling et al., 2010).

In the present study, the genomes and metabolomes of *in situ* samples from three Austrian natural populations of *A. thaliana* were characterized. Applying a combination

<sup>1</sup><http://www.1001genomes.org>



of metabolomics, multivariate statistics, and mathematical modeling based on genome-derived biochemical pathway information, biochemical and physiological signatures of *in situ* *Arabidopsis* populations could be identified (Figure 1). Different metabolic steady states on a population level and general patterns common to all populations were distinguished by this integrative approach, which finally allowed the prediction of characteristic processes of *in situ* metabolic adaptation.

## MATERIALS AND METHODS

### Plant Material and Sampling Strategy

*In situ* sampling of *A. thaliana* leaf rosettes was performed in three Austrian locations (see Figure 2). The first location (OOE1) was a hay meadow, the second (OOE2) was a rocky spot with variable substrate thickness, and the third sampling site (OOE3) was an unused meadow with steep slope and a nearby valley. All populations were located in close proximity to intensively used grassland. Each sample consisted of one whole leaf rosette without inflorescence. Global positioning system (GPS) coordinates of the sampling sites were recorded using a Garmin Oregon300 handheld GPS receiver (Garmin®, Schaffhausen, Switzerland) with an accuracy of approximately 3 m. The waypoints were imported into Garmin Mapsource Version 6.15.6 (Garmin®, Schaffhausen, Switzerland) and projected on the OpenStreetMap<sup>2</sup>. The

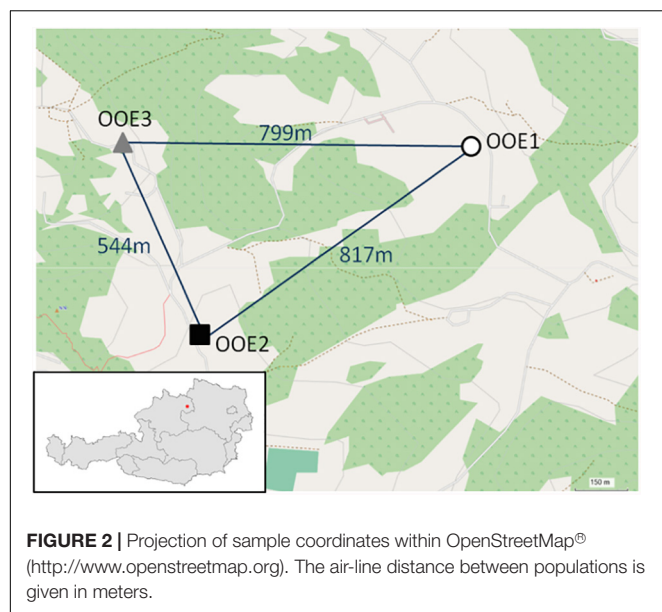
sampling was performed according to Scherling et al. (2010) with a minimized cycle of time accounting for diurnal changes. The sampling began at 12 am at OOE1, then OOE2 and OOE3. The sampling was repeated three times at the same day comprising about 20 min each and was finished at 4 pm. The sampling day had continuous cloudy and constant weather conditions. All *Arabidopsis* rosettes were sampled at a developmental stage in which inflorescence and mature leaf rosettes had been established (example pictures are provided in Supplementary Data Sheet S1). Altogether we sampled  $n = 13$ , 15, and 15 biological replicates for OOE1, OOE2, and OOE3, respectively, for GC-MS and  $n = 10$ , 7, and 13 biological replicates for OOE1, OOE2, and OOE3, respectively, for LC-MS analyses. Rosettes were cut and immediately frozen in liquid nitrogen. Samples were stored at  $-80^{\circ}\text{C}$  until further processing.

### DNA Sequencing and SNP Calling

Sequencing was performed for individual plants of the different populations OOE1, OOE2 and OOE3. Genomic DNA preparation, and SNP calling was performed as described previously (Alonso-Blanco et al., 2016). The samples were sequenced using 100 bp paired-end reads on an Illumina HiSeq platform. Pairwise genetic differences ( $\theta_p$ ) between these accessions and a set of additional 24 accessions for which DNA sequence is publically available (see footnote 1) has been calculated by dividing the number of polymorphic sites by the number of informative sites. These values have been used to create a hierarchical clustering using the McQuitty

<sup>2</sup><http://www.freizeitkarte-osm.de/de/oesterreich.html>





method within the function *hclust* in R (McQuitty, 1966)<sup>3</sup>. To extract the most diverse genes from the three populations, we calculated the amount of variation between the populations for each gene. We used only sides where we had SNP calls for one representative of each population. We created a list for each population containing only genes that differ by at least 50 polymorphisms from the other two populations. These lists are available as **Supplementary Data Sheets S2–S4**. Furthermore, we created population-specific clustered protein interaction networks with these genes using STRING (Szklarczyk et al., 2017). In **Supplementary Presentation S1** the networks and gene functions are shown. All SNP data are stored at the public repository (see footnote 1).

## Gas Chromatography Coupled to Time-of-Flight Mass Spectrometry

Frozen sample rosettes were homogenized in a ball mill (Retsch®, Haan, Germany) under frequent cooling with liquid nitrogen for 3 min. Polar metabolites were extracted and derivatized as described previously (Weckwerth et al., 2004b). Gas chromatography coupled to mass spectrometry (GC-MS) analysis was performed on an Agilent 6890 gas chromatograph (Agilent Technologies®, Santa Clara, CA, United States) coupled to a LECO Pegasus® 4D GCxGC-TOF mass spectrometer (LECO Corporation, St. Joseph, MI, United States). Compounds were separated on an Agilent HP5MS column (length: 30 m length, diameter: 0.25 mm, film: 0.25 µm). Deconvolution of the total ion chromatograms was performed using the LECO Chromatof® software. All details about injection, gradient, deconvolution, and library search parameters can be found in Doerfler et al. (2013). A calibration curve was recorded for absolute quantification of central primary metabolites.

<sup>3</sup><https://www.rdocumentation.org/packages/stats/versions/3.5.1/topics/hclust>

## GC-MS Data Analysis and Inverse Approximation of Jacobian Matrix Entries

For ANOVA and computation of *p*-values adjusted for sample size by Tukey Honest Significant Differences R was used (R, 2013). For multivariate analysis, outliers (all values that were lower/higher than 1.5\*interquartile range from the 25%/75% quantile) were removed from the dataset. Missing values of variables, which were missing in more than half of all measurements in a population were filled with half of the matrix minimum. The remaining missing values were imputed by random forest computation (Stekhoven and Buhlmann, 2012; Gromski et al., 2014). This dataset was centered and scaled to unit variance prior to sPLS regression. Sparse partial least squares (sPLS) regression analysis was performed using the mixOmics package (Le Cao et al., 2009; Gonzalez et al., 2011, 2012) for the statistical software environment R (R, 2013).

The functional integration of GC-MS metabolomics data into a metabolic network was performed, as previously described (Nägele et al., 2014), by the approximation of the biochemical Jacobian matrix. This approximation directly connects the covariance matrix *C*, which was built from the experimental metabolomics data, with a metabolic network structure derived from *Arabidopsis* genome information. Linkage of covariance data with the network structure follows equation 1 (Steuer et al., 2003; Sun and Weckwerth, 2012):

$$JC + CJ^T = -2D \quad (1)$$

Here, *J* represents the Jacobian matrix and *D* is a fluctuation matrix which integrates a Gaussian noise function simulating metabolic fluctuations around a steady state condition. In context of a metabolic network, entries of the Jacobian matrix *J* represent the elasticity of reaction rates to any change of metabolite concentrations which are characterized by equation 2:

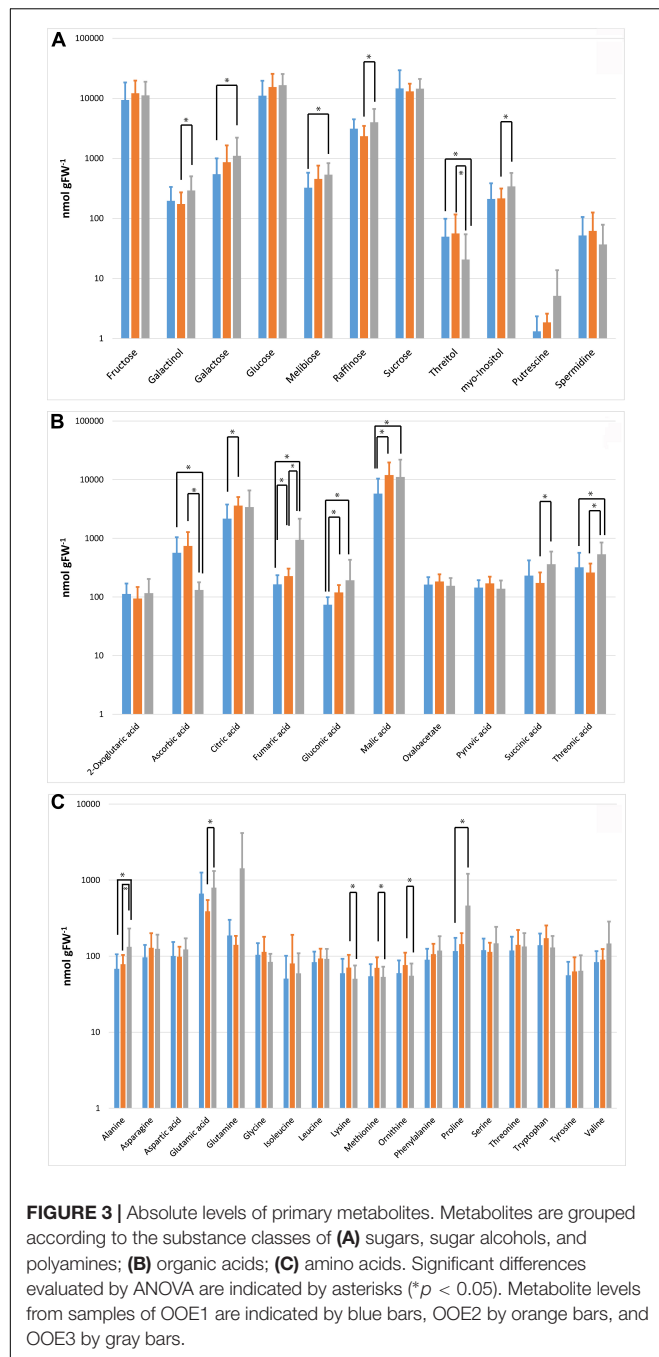
$$J = N \frac{\partial r}{\partial M} \quad (2)$$

*N* is the stoichiometric matrix or a metabolic interaction matrix if reactions and reactants have been modified in the original network. *r* represents the rates for each reaction, and *M* represents metabolite concentrations. As stated before, the Jacobian approximation comprises the stochastic term *D*. Therefore, we performed  $10 \times 10^5$  inverse approximations for each population, finally resulting in 10 technical replicates of the Jacobian matrices.

All calculations of Jacobian matrices were performed based on a modified version of the toolbox COVAIN (Sun and Weckwerth, 2012) within the numerical software environment MATLAB® (V8.4.0 R2014b).

## LC-MS Analysis

The frozen plant leaf material was homogenized and extracted as the samples for the GC-MS analysis as described recently (Weckwerth et al., 2004b; Doerfler et al., 2013). The polar fraction of metabolites was dried in a speedvac. Extracts were weighed



and dissolved in 5% Acetonitrile 0.5% Formic acid to an extract concentration of 0.5 g/L. From these solutions, 3  $\mu$ L were injected to an Agilent Ultimate 3000 LC-system and separated on a reversed-phase column on a 60-min effective gradient prior to data-dependent mass spectrometric analysis of +1 – charged ions (Doerfler et al., 2013, 2014). Acquired LC-MS runs were converted to the open mzXML data format using the MassMatrix File Conversion Tools. Subsequently, MS1 intensities of all mass traces that were fragmented at least once in a sample were summed over the whole runs with ProtMAX2012 (Hohenwarter

et al., 2011; Doerfler et al., 2013; Egelhofer et al., 2013). The data set was filtered for features that were measured in at least half of the replicates of one population and remaining variables were normalized to the sum of all variables of the respective sample. The resulting values were used to fit ANOVA models. Tukey Honest Significant Differences were used to estimate sample-size adjusted  $p$ -values in R (R, 2013). Venny was used to visualize the number of detected significant differences (Oliveros, 2007).

For multivariate analysis, outliers (all values that were lower/higher than 1.5\*interquartile range from the 25%/75% quantile) were removed from the dataset. Missing values of variables, which were missing in more than half of all measurements in a population were filled with half of the matrix minimum. The remaining missing values were imputed by random forest computation (Stekhoven and Buhlmann, 2012; Gromski et al., 2014). This dataset was centered and scaled to unit variance prior to sPLS regression (see above).

## RESULTS

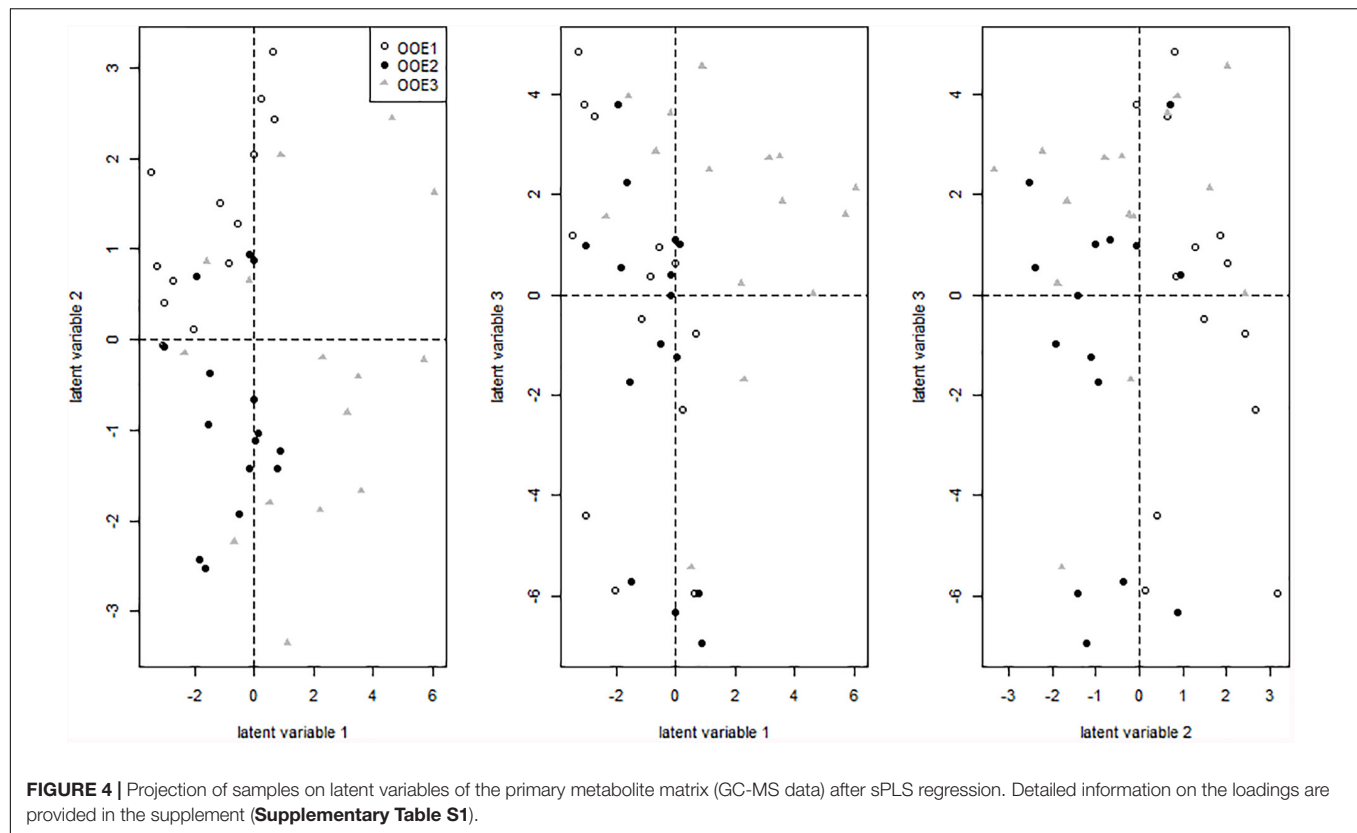
### Metabolomic Analysis of *in situ* Samples

*In situ* sampling of *A. thaliana* leaf rosettes was performed on three nearby locations in Upper Austria (Oberösterreich; OOE; see Figure 2 and See section “Materials and Methods”). All *Arabidopsis* rosettes were sampled at a developmental stage in which inflorescence and mature leaf rosettes had been established (example pictures are provided in **Supplementary Data Sheet S1**). For a set of metabolites from untargeted GC-MS based metabolomics data, we performed absolute quantification using calibration curves. This set of metabolites comprised concentrations of 39 central compounds of the C/N metabolism including sugars and sugar alcohols, organic acids, amino acids, and polyamines (Figure 3). Results of an ANOVA indicated that only levels of fumaric acid discriminated all three populations significantly (Figure 3B). Populations OOE1 and OOE3 could be discriminated significantly by the concentrations of galactose, melibiose, threitol, ascorbic acid, fumaric acid, gluconic acid, malic acid, threonic acid, alanine, and proline ( $p < 0.05$ ; Figure 3). For populations OOE2 and OOE3, significant differences were found to exist for absolute levels of galactinol, raffinose, threitol, myo-inositol, ascorbic acid, fumaric acid, succinic, and threonic acid as well as for the amino acids alanine, glutamic acid, lysine, methionine, and ornithine ( $p < 0.05$ ; Figure 3). Populations OOE1 and OOE2 could be discriminated by levels of citric acid, fumaric acid, gluconic acid, and malic acid. To summarize these findings, most significant differences between absolute metabolite levels of populations OOE1, 2, and 3 were determined for the class of organic acids (13 out of 27, i.e., ~50%).

### Multivariate Analysis Indicates a Discrimination of *in situ* Populations by Metabolic Phenotypes

Sparse partial least squares (sPLS) regression analysis of primary metabolites versus a response matrix comprising geographical



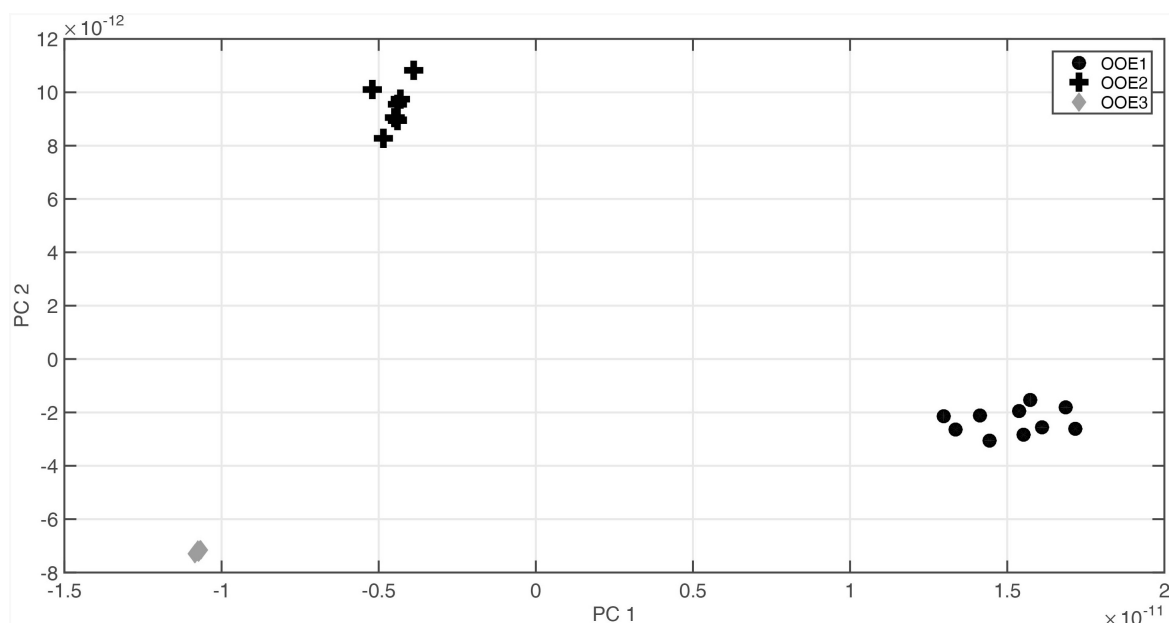


coordinates and altitude above sea level indicated a separation of population OOE3 from populations OOE1 and OOE2 across latent variable 1 (**Figure 4**). The metabolite levels of fumaric acid, melibiose, alanine, putrescine, gluconic acid, threonic acid, myo-inositol, galactinol and succinic acid were identified to contribute most to this separation with elevated levels in OOE3 whereas mainly ascorbic acid and threitol were elevated in OOE1 and OOE2. Discrimination of populations OOE1 and OOE2 was indicated on latent variable 2 (**Figure 4**). Here, a higher abundance of 2-oxoglutaric acid, glutamic acid, raffinose, glycine, succinic acid, serine and threonic acid in OOE1 and malic acid, gluconic acid and citric acid in OOE2 was observed (see **Supplementary Table S1** for a complete list of loadings, table sheet “Loadings GCMS”, and **Supplementary Figure S1** for a PCA analysis of the primary metabolites).

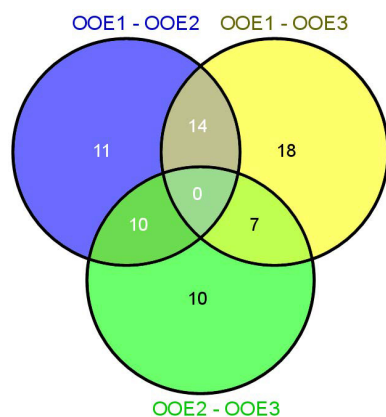
### Entries of Jacobian Matrices Indicate Different Biochemical Phenotypes of the *in situ* Populations at the Interface of Primary and Secondary Metabolism

While absolute metabolite levels can provide a representative view on a metabolic homeostasis, it can hardly be directly interpreted in terms of biochemical regulation. Instead, strategies of multivariate statistics and modeling were shown to be essential to provide a comprehensive view on the biochemical regulation of a metabolic homeostasis (Weckwerth, 2011b). Based on a biomathematical strategy developed and applied in former

studies, entries of Jacobian matrices were directly inferred from experimental metabolomic covariance data (Doerfler et al., 2013; Nägele et al., 2014) (**Figure 1**). As described in our previous work, we derived a metabolic network model comprising reactants and reactions indicated in the **Supplementary Table S2**. The metabolic covariance information was linked to a genome-information derived biochemical network structure, finally satisfying a Lyapunov matrix equation [for more details about the method and the metabolic network model, we refer to the section Materials and Methods as well as to our previous work (Nägele et al., 2014)]. The calculation procedure, that is, solving the equation after stochastic perturbation, was performed  $10 \times 10^5$  times and median values of all entries of the Jacobian matrices were determined. Principal component analysis (PCA) of the entries revealed a clear separation of the population-specific Jacobian information in which the technical variance was found to be significantly lower than the biological variance (**Figure 5**). Loadings of the PCA revealed that the strong separation of population OOE1 from OOE2 and 3 on component 1 (PC1) was predominantly due to differences in organic and amino acid, polyamine, and raffinose metabolism but also aromatic amino acid biosynthesis and interconversion (**Supplementary Tables S2, S3**). This output indicated a potential difference in the regulation of secondary metabolism, or, at least, at the interface between primary and secondary metabolism. Hence, secondary metabolite abundance of the three Austrian *Arabidopsis* populations was recorded applying LC-MS analysis. The quantitative analysis of specific mass traces in the



**FIGURE 5 |** Principal component analysis (PCA) of Jacobian matrix entries for populations OOE1, OOE2, and OOE3. PC1 strongly separates OOE1 (black filled circles) from OOE2 (black filled crosses) and OOE3 (gray filled diamonds). PC2 separates OOE2 from OOE1 and OOE3 most significantly. Detailed information about the Jacobian entries and the loadings can be found in **Supplementary Tables S2, S3**.



**FIGURE 6 |** Venn diagram showing the number of LC-MS features, which significantly discriminated the three *Arabidopsis* populations OOE1, 2, and 3.

chromatograms showed that there was no feature separating all of the populations significantly (ANOVA,  $p < 0.05$ ). Yet, we were able to identify 70 features that discriminated at least two of the populations (Figure 6).

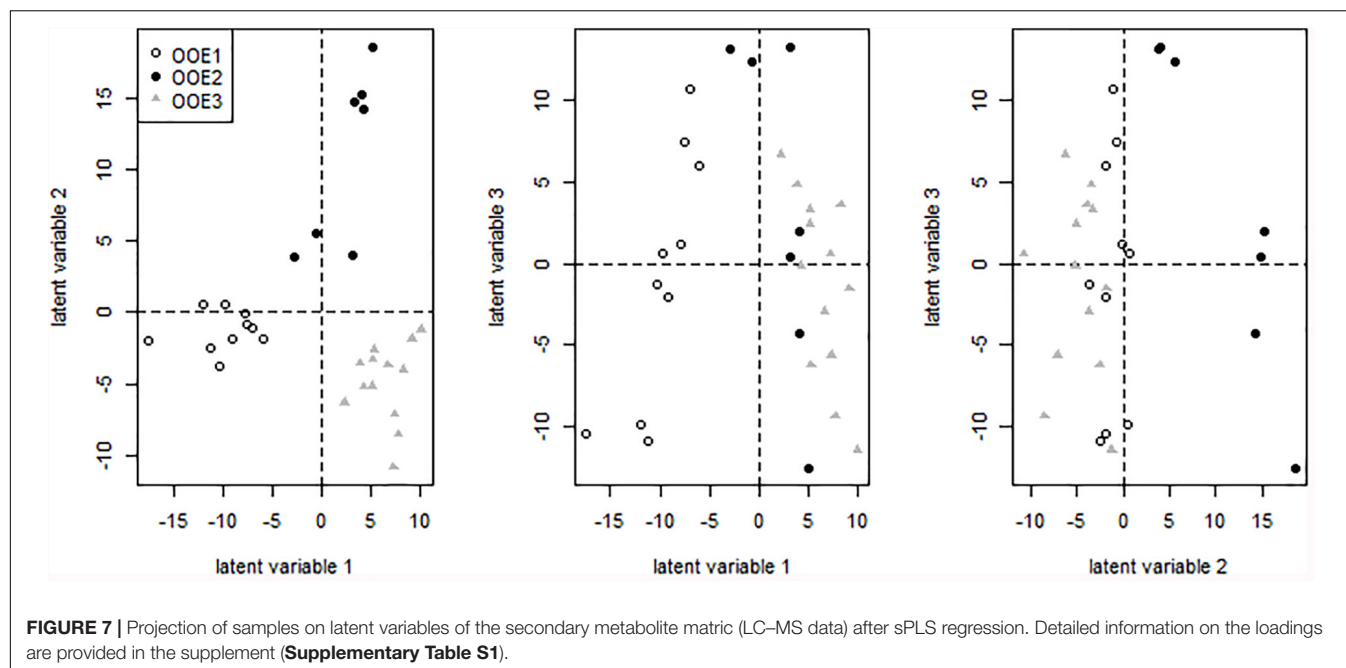
To statistically evaluate the separation of populations by secondary metabolites, LC-MS data were analyzed by sPLS regression analysis. The first latent variable was found to separate OOE1 from OOE2 and OOE3 (Figure 7; Loadings are provided in **Supplementary Table S1**, table sheet “Loadings LCMS”). The second latent variable indicated a separating effect of several putative anthocyanins attached to sinapoyl moieties

[A6, A7/A17, A8, A10, A11, and  $m/z$  1329, respectively, for further annotation see Doerfler et al. (2014)] in the OOE2 population by which it was discriminated from OOE1 and OOE3.

## Genotyping of *in situ* Natural *A. thaliana* Populations

A SNP-based genotyping approach was performed to unravel the genomic relationship of the three populations. Genotyping showed clear differences between the three populations (Figure 8). Different individual plants of population OOE2 were found to be nearly identical (12, 23, and 13 SNPs, respectively). The population OOE2 was found to differ by approximately 300,000 SNPs from both populations OOE1 and OOE3, which were likewise separated by more than 300,000 SNPs. Interestingly, individual plants that have been sequenced from the OOE3 population were genetically different as well but to a minor extent ( $\sim 260,000$  SNPs). The comparison with genomic data from other ecotypes show the expected genetic differences not only between these populations but also with respect to global samples, in which accessions from Austria, Italy, and the Czech Republic are most similar (Figure 8). To extract the most diverse genes from the three populations, we created a list for each population containing only genes that differ by at least 50 polymorphisms from the other two populations. These lists are available as **Supplementary Data Sheets S2–S4**. Furthermore, we created population-specific clustered protein interaction networks with these genes using STRING (Szklarczyk et al., 2017). These protein interaction networks showed highly diverse functional pattern





between the three different populations (see **Supplementary Presentation S1**).

## DISCUSSION

### Eco-Metabolomics and Metabolic Modeling

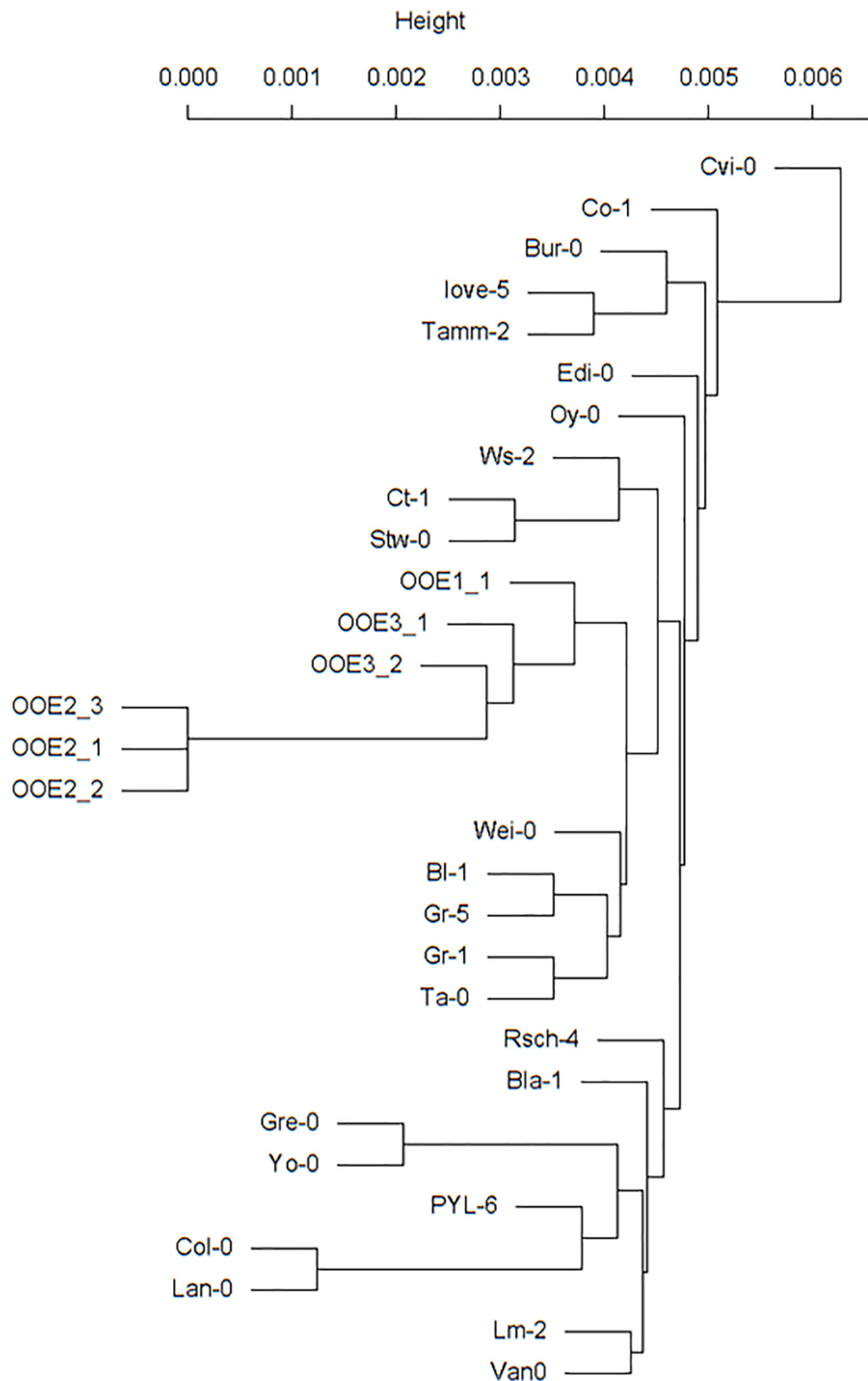
The importance and central role of metabolomics in an ecological context has extensively been outlined in previous studies and overview articles [see e.g., Sardans et al., 2011; Jones et al., 2013]. One of the central issues of eco-metabolomic approaches is the detection and characterization of environmentally induced phenotypic mechanisms in context of key metabolic processes and ecologically relevant parameters, that is, all kinds of environmental cues (Scherling et al., 2010). Yet, due to the non-linear relationship between single levels of molecular organization, the reliable interpretation of metabolomics results is highly challenging. The metabolic output or homeostasis of a biochemical system depends on numerous molecular parameters and variables, and, finally, a metabolic network sums up to a highly branched, interlaced and non-linearly behaving molecular system (Nägele, 2014).

While under controlled conditions such plasticity of molecular systems already significantly limits our ability to intuitively draw conclusions about regulatory mechanisms, *in situ* data interpretation has to cope even more with a potential ambiguousness introduced by environmental dynamics and fluctuations (Macel et al., 2010). In the present study, such fluctuations were taken into account by considering (co)variance information of metabolite pools and by a modeling approach, which focuses on the characterization of dynamical behavior of metabolic systems around a metabolic homeostasis (Nägele

et al., 2014) (see **Figure 1**). In detail, data dimensionality reduction via PCA indicated a clear separation of all populations by Jacobian entries being related to the biochemistry at the interface of primary and secondary metabolism as well as the metabolism of metabolic stress-markers, such as polyamines and raffinose, which have been discussed to be involved in the protection of the photosynthetic apparatus against various stress types (Bouchereau et al., 1999; Alcázar et al., 2006, 2011; Knaupp et al., 2011).

### Plasticity of Plant Primary Metabolism in *in situ* Populations and Correlation With Geographical Coordinates

Statistics on absolute primary metabolite levels revealed major differences between natural *in situ* Austrian *A. thaliana* populations. Almost all classes of analyzed substances, comprising sugars, carboxylic, and amino acids displayed significant differences indicating different homeostasis in primary metabolism of all three populations. The TCA intermediate fumaric acid was found to significantly differ between all *in situ* samples indicating suitability to classify these populations. While it has been shown that fumaric acid metabolism plays a central role in diurnal carbon allocation (Pracharoenwattana et al., 2010), and, hence, indirectly affects the orchestration of photosynthesis in *Arabidopsis* leaves, it remains to be demonstrated whether it can directly report on changes in plant–environment interactions. In addition, due to the complex regulation of plant primary metabolism, it can hardly be assumed that one metabolite level provides representative information for robust metabolic *in situ* classification. Yet, together with the finding of a significant difference in potential photosystem-protective substances, for example, polyamines and flavonoids, it can be hypothesized that differential metabolic homeostasis



**FIGURE 8 |** SNP genotyping of three Austrian *Arabidopsis* populations. All three plants that have been sequenced from the population OOE2 are nearly genetically identical. OOE2 differs by nearly 300,000 SNPs from both the OOE1 and OOE3 population, which are likewise separated by more than 300,000 SNPs. The comparison with genomic data from other ecotypes showed the expected genetic differences not only within these populations but also to global samples, in which accessions from Austria, Italy, and the Czech Republic are most related. The genome information of all accessions is publically available at [www.1001genomes.org](http://www.1001genomes.org).

evolved due to differences in the microenvironment of the three populations being characteristic enough to separate them according to the resulting metabolic signatures. We further asked the question whether we can identify metabolite marker, which show significant correlations with geographical coordinates even within this proximate distribution of populations. In **Supplementary Presentation S2** a correlation network is shown between geographical coordinates and primary metabolites of the three different populations. Indeed, there is a clear distinction between several metabolites showing significant correlations to altitude, E and N coordinates. This further provides evidence that metabolic homeostasis is related to environmental differences between these different locations of the natural populations.

## The Interface of Primary and Secondary Metabolism as a Key Regulatory Point for Genotypic and Phenotypic Plasticity

Predictions about the differentiation via signatures in secondary metabolism were validated by LC-MS metabolomics focusing on a central set of secondary metabolite backbones with close similarity to previously identified anthocyanins attached to sinapoyl moieties (Doerfler et al., 2014). Such metabolic differences are in line with previous findings reporting on metabolic signatures, which are due to characteristic differences in specialized or secondary metabolism (Wink, 2003; Lu et al., 2009; Scherling et al., 2010; Doerfler et al., 2013; Chae et al., 2014; Moore et al., 2014). The accumulation of anthocyanin pigments in vegetative tissue of plants represents an approved metabolic stress and acclimation output (Winkel-Shirley, 2002). Moreover, we demonstrated earlier that most of the statistical significant metabolic responses of individual plant species to *in situ* biodiversity are attributed to secondary metabolism including flavonoid structures (Scherling et al., 2010). Hence, the molecular analysis provided a detailed view on the differential population-specific metabolic composition of secondary metabolites and anthocyanin-related leaf color. With this, evidence is provided for the suitability of metabolic phenotyping of *in situ* samples by a combined GC-MS and LC-MS platform (Scherling et al., 2010; Doerfler et al., 2013). While, at this point, we can only speculate on the environmental cues which initiated the observed differences in secondary stress-associated metabolism, flavonoid metabolites in general are heavily discussed in context of their UV absorption and reactive oxygen species (ROS) scavenging properties (Winkel-Shirley, 2002; Agati and Tattini, 2010; Doerfler et al., 2014; Hectors et al., 2014). Together with the finding of a differentially regulated polyamine metabolism between the populations, which became visible rather by covariance information than by mean values, our results point toward a differential macro- or microclimatic environment at the three Austrian *in situ* sampling sites (see also description of the sampling sites in Materials and Methods).

In addition, results of SNP-based genotyping revealed three genetically different populations, which are, however, closer related to each other than to other European accessions

(**Figure 8**). In terms of temperature regimes and low temperature tolerance, which can be expected to have major influence on the geographic range of *A. thaliana* (Hoffmann, 2002), the genetic distance between the Austrian populations can be regarded as relatively small when compared to sensitive (Cvi, Co-1) and tolerant accessions (Rsch-4) (Hannah et al., 2006). Based on this observation, we hypothesize that the variance in observed metabolic phenotypes are a mixture of plasticity effects and conceptual differences in the acquisition of abiotic stress tolerance. This again might indicate a high intraspecific metabolic variation, which would affect the evolutionary capacity of *Arabidopsis* in context of adaptation to macro- and micro-environmental fluctuation (Moore et al., 2014).

## Jacobian Entries Are Potentially Linked to Intraspecific Genotypic Variation

The combination of in depth genotypic and metabolomic profiling and modeling opens up the opportunity to search for direct correlations of polymorphisms and metabolic changes. Here, we applied this concept to an *in situ* study for the first time and revealed a significant intraspecific biochemical plasticity within three close-by natural populations in their natural habitat. We have extracted the genes of the individual populations which distinguish them most (**Supplementary Data Sheets S2–S4**). By further analysis of the corresponding clustered protein interaction networks different functional modules between the different populations became visible (**Supplementary Presentation S1**). The three populations OOE1, 2, and 3 showed severe differences in these protein interaction networks. Especially, the OOE3 population showed a cluster of genes which is clearly involved in organic acid and amino acid metabolism including genes for pyruvate dehydrogenase, aconitase, NAD-malic enzyme 1, pyruvate-phosphate dikinase, lactate-dehydrogenase, and several others (see **Supplementary Presentation S1**). These functional patterns, which distinguish OOE3 from OOE1 and 2 coincide with the calculation of Jacobian entries. The strongest loadings separating OOE3, 2, and 1 on PC1 in **Figure 4** include  $df(\text{Glu})/d(\text{Pyr})$ ,  $df(\text{Mal})/d(\text{Fum})$ ,  $df(\text{Cit})/d(\text{Pyr})$ ,  $df(\text{Glu})/d(\text{Asp})$ ,  $df(\text{Glu})/d(2\text{-oxoglutarate})$ ,  $df(\text{Succ})/d(\text{Put})$ . All of these entries point to organic acid metabolism and key entry points for amino acid metabolism, especially nitrogen assimilation and transamination reactions. In future studies, we will investigate these relationships in more detail also by integrating proteomics studies. There is a great potential that the calculation of Jacobian entries of a biochemical matrix gives important clues about different dynamics in the same set of metabolites based on intraspecific but also interspecific genetic variance and biochemical regulation. This is due to the explicit linkage of the metabolite covariance matrix  $C$  – representing the dynamic part of the equation – and the Jacobian  $J$ , which relies on the metabolite interaction matrix defined by genome-scale metabolic reconstruction and biochemical pathways. Accordingly, the covariance matrix  $C$  is representative for the different ecotype dynamics whereas the Jacobian structure preserves the reconstructed metabolic network. Just the combination of both  $J$  and  $C$  in the



Lyapunov matrix equation will reveal the dynamics of each ecotype individually (for further details see also Weckwerth, 2011a,b).

## CONCLUSION

In summary, it was demonstrated that intraspecific metabolic phenotypes of geographically nearby-grown *Arabidopsis* plants can be characterized and differentiated by their primary-secondary metabolic signature. In future studies, monitoring of micro-climatic properties will enable the characterization of sampling sites by continuous quantitative environmental data and thus improve the understanding of the ecological context of *in situ* molecular profiles. Additionally, biotic and abiotic habitat parameters, such as soil properties and phytosociological association, might even promote our current understanding of individual plants' physiology. Finally, our study points to the importance of considering variance and covariance information in biological data sets (Weckwerth et al., 2004b; Violle et al., 2012) which, together with genome-derived pathway information, potentially provide information about environmental fluctuations, and associated biochemical system properties. The findings contribute to the comprehensive understanding of ecological processes and may contribute to the design of future studies focusing on the estimation of the impact of climate change on plant societies and evolution using combined multiomics and modeling strategies (Ward and Kelly, 2004; Weckwerth, 2011a).

## AUTHOR CONTRIBUTIONS

MNa collected natural populations of *Arabidopsis thaliana*, performed the measurements and statistical analysis, and wrote the manuscript. TN performed the measurements, statistical analysis, metabolic modeling, and wrote the manuscript. CG identified and collected natural populations of *Arabidopsis thaliana*. LF harvested sample material. AK, AP, AE, and MNo performed SNP calling, population analysis, and wrote the

manuscript. WW conceived the study, performed statistical analysis, and wrote the manuscript.

## FUNDING

The study was funded by the Austrian Science Fund (FWF, Project P 26342, and I 1022).

## ACKNOWLEDGMENTS

The authors would like to thank the members of the Department Ecogenomics and Systems Biology for valuable discussions and advice. Additionally, they thank the gardeners and the whole team of the department-associated greenhouse facility for their support and advice.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.01556/full#supplementary-material>

**FIGURE S1** | PCA analysis of primary metabolites.

**TABLE S1** | PCA loadings of GC-MS and LC-MS metabolites.

**TABLE S2** | Table of Jacobian entries and their associated metabolite, pathway and enzyme reaction (EC number).

**TABLE S3** | PCA loadings of the Jacobien entries.

**DATA SHEET S1** | Examples of individual plants of the three *Arabidopsis thaliana* populations OOE1-3.

**DATA SHEET S2** | SNP enriched genes in OOE1.

**DATA SHEET S3** | SNP enriched genes in OOE2.

**DATA SHEET S4** | SNP enriched genes in OOE3.

**PRESENTATION S1** | String protein interaction networks for SNP enriched genes distinguishing the three natural *Arabidopsis thaliana* populations OOE1-3.

**PRESENTATION S2** | Metabolite – GPS – correlation network.

## REFERENCES

- Agati, G., and Tattini, M. (2010). Multiple functional roles of flavonoids in photoprotection. *New Phytol.* 186, 786–793. doi: 10.1111/j.1469-8137.2010.03269.x
- Alcázar, R., Cuevas, J. C., Planas, J., Zarza, X., Bortolotti, C., Carrasco, P., et al. (2011). Integration of polyamines in the cold acclimation response. *Plant Sci.* 180, 31–38. doi: 10.1016/j.plantsci.2010.07.022
- Alcázar, R., Marco, F., Cuevas, J. C., Patron, M., Ferrando, A., Carrasco, P., et al. (2006). Involvement of polyamines in plant response to abiotic stress. *Biotechnol. Lett.* 28, 1867–1876. doi: 10.1007/s10529-006-9179-3
- Alonso-Blanco, C., Andrade, J., Becker, C., Bemm, F., Bergelson, J., Borgwardt, K. M., et al. (2016). 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell* 166, 481–491. doi: 10.1016/j.cell.2016.05.063
- Arabidopsis Genome Initiative [AGI] (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815. doi: 10.1038/35048692
- Atwell, S., Huang, Y. S., Vilhjalmsón, B. J., Willems, G., Horton, M., Li, Y., et al. (2010). Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465, 627–631. doi: 10.1038/nature08800
- Baerenfaller, K., Bastow, R., Beynon, J., Brady, S., Brendel, V., Donaldson, S., et al. (2012). Taking the next step: building an *Arabidopsis* information portal. *Plant Cell* 24, 2248–2256. doi: 10.1105/tpc.112.100669
- Bouchereau, A., Aziz, A., Larher, F., and Martin-Tanguy, J. (1999). Polyamines and environmental challenges: recent development. *Plant Sci.* 140, 103–125. doi: 10.1016/S0168-9452(98)00218-0
- Chae, L., Kim, T., Nilo-Poyanco, R., and Rhee, S. Y. (2014). Genomic signatures of specialized metabolism in plants. *Science* 344, 510–513. doi: 10.1126/science.1252076
- Chan, E. K., Rowe, H. C., Hansen, B. G., and Kliebenstein, D. J. (2010). The complex genetic architecture of the metabolome. *PLoS Genet.* 6:e1001198. doi: 10.1371/journal.pgen.1001198
- Darwin, C. (1859). *The Origin of Species*. London: John Murray.

- Diz, A. P., Martínez-Fernández, M., and Rolán-Alvarez, E. (2012). Proteomics in evolutionary ecology: linking the genotype with the phenotype. *Mol. Ecol.* 21, 1060–1080. doi: 10.1111/j.1365-294X.2011.05426.x
- Doerfler, H., Lyon, D., Nagele, T., Sun, X., Fragner, L., Hadacek, F., et al. (2013). Granger causality in integrated GC-MS and LC-MS metabolomics data reveals the interface of primary and secondary metabolism. *Metabolomics* 9, 564–574. doi: 10.1007/s11306-012-0470-0
- Doerfler, H., Sun, X., Wang, L., Engelmeier, D., Lyon, D., and Weckwerth, W. (2014). mzGroupAnalyzer—predicting pathways and novel chemical structures from untargeted high-throughput metabolomics data. *PLoS One* 9:e96188. doi: 10.1371/journal.pone.0096188
- Egelhofer, V., Hoehenwarter, W., Lyon, D., Weckwerth, W., and Wienkoop, S. (2013). Using ProtMAX to create high-mass-accuracy precursor alignments from label-free quantitative mass spectrometry data generated in shotgun proteomics experiments. *Nat. Protoc.* 8, 595–601. doi: 10.1038/nprot.2013.013
- Ellenberg, H. (1953). Physiologisches und ökologisches Verhalten derselben pflanzenarten. *Ber. Deutsch. Bot. Ges.* 65, 351–362.
- Gonzalez, I., Cao, K. A., Davis, M. J., and Dejean, S. (2012). Visualising associations between paired 'omics' data sets. *Biodata Min.* 5:19. doi: 10.1186/1756-0381-5-19
- Gonzalez, I., Lê Cao, K., and Déjean, S. (2011). *MixOmics: Omics Data Integration Project*. Available at: <http://www.math.univ-toulouse.fr/~biostat/mixOmics>
- Gromski, P. S., Xu, Y., Kotze, H. L., Correa, E., Ellis, D. I., Armitage, E. G., et al. (2014). Influence of missing values substitutes on multivariate analysis of metabolomics data. *Metabolites* 4, 433–452. doi: 10.3390/metabo4020433
- Hancock, A. M., Brachi, B., Faure, N., Horton, M. W., Jarymowycz, L. B., Sperone, F. G., et al. (2011). Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* 334, 83–86. doi: 10.1126/science.1209244
- Hannah, M. A., Wiese, D., Freund, S., Fiehn, O., Heyer, A. G., and Hinch, D. K. (2006). Natural genetic variation of freezing tolerance in *Arabidopsis*. *Plant Physiol.* 142, 98–112. doi: 10.1104/pp.106.081141
- Hector, A., Von Felten, S., Hautier, Y., Weilenmann, M., and Bruelheide, H. (2012). Effects of dominance and diversity on productivity along ellenberg's experimental water table gradients. *PLoS One* 7:e43358. doi: 10.1371/journal.pone.0043358
- Hectors, K., Van Oevelen, S., Geuns, J., Guisez, Y., Jansen, M. A. K., and Prinsen, E. (2014). Dynamic changes in plant secondary metabolites during UV acclimation in *Arabidopsis thaliana*. *Physiol. Plant* 152, 219–230. doi: 10.1111/ppl.12168
- Hoehenwarter, W., Larhlumi, A., Hummel, J., Egelhofer, V., Selbig, J., Van Dongen, J. T., et al. (2011). MAPA distinguishes genotype-specific variability of highly similar regulatory protein isoforms in potato tuber. *J. Proteome Res.* 10, 2979–2991. doi: 10.1021/pr101109a
- Hoehenwarter, W., Van Dongen, J. T., Wienkoop, S., Steinfath, M., Hummel, J., Erban, A., et al. (2008). A rapid approach for phenotype-screening and database independent detection of cSNP/protein polymorphism using mass accuracy precursor alignment. *Proteomics* 8, 4214–4225. doi: 10.1002/pmic.200701047
- Hoffmann, M. H. (2002). Biogeography of *Arabidopsis thaliana* (L.) Heynh. (*Brassicaceae*). *J. Biogeogr.* 29, 125–134. doi: 10.1046/j.1365-2699.2002.00647.x
- Horton, M. W., Hancock, A. M., Huang, Y. S., Toomajian, C., Atwell, S., Auton, A., et al. (2012). Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions from the RegMap panel. *Nat. Genet.* 44, 212–216. doi: 10.1038/ng.1042
- Hunter, B., Wright, K. M., and Bomblies, K. (2013). Short read sequencing in studies of natural variation and adaptation. *Curr. Opin. Plant Biol.* 16, 85–91. doi: 10.1016/j.pbi.2012.10.003
- Jones, O. A. H., Maguire, M. L., Griffin, J. L., Dias, D. A., Spurgeon, D. J., and Svendsen, C. (2013). Metabolomics and its use in ecology. *Aust. Ecol.* 38, 713–720. doi: 10.1111/aec.12019
- Keurentjes, J. J. B. (2009). Genetical metabolomics: closing in on phenotypes. *Curr. Opin. Plant Biol.* 12, 223–230. doi: 10.1016/j.pbi.2008.12.003
- Kleessen, S., Antonio, C., Sulpice, R., Laitinen, R., Fernie, A. R., Stitt, M., et al. (2012). Structured patterns in geographic variability of metabolic phenotypes in *Arabidopsis thaliana*. *Nat. Commun.* 3:1319. doi: 10.1038/ncomms2333
- Knaupp, M., Mishra, K. B., Nedbal, L., and Heyer, A. G. (2011). Evidence for a role of raffinose in stabilizing photosystem II during freeze-thaw cycles. *Planta* 234, 477–486. doi: 10.1007/s00425-011-1413-0
- Koornneef, M., Alonso-Blanco, C., and Vreugdenhil, D. (2004). Naturally occurring genetic variation in *Arabidopsis thaliana*. *Annu. Rev. Plant Biol.* 55, 141–172. doi: 10.1146/annurev.arplant.55.031903.141605
- Lasky, J. R., Des Marais, D. L., Mckay, J. K., Richards, J. H., Juenger, T. E., and Keitt, T. H. (2012). Characterizing genomic variation of *Arabidopsis thaliana*: the roles of geography and climate. *Mol. Ecol.* 21, 5512–5529. doi: 10.1111/j.1365-294X.2012.05709.x
- Lavagi, I., Estelle, M., Weckwerth, W., Beynon, J., and Bastow, R. M. (2012). From bench to bountiful harvests: a road map for the next decade of *Arabidopsis* research. *Plant Cell* 24, 2240–2247. doi: 10.1105/tpc.112.096982
- Le Cao, K. A., Gonzalez, I., and Dejean, S. (2009). integrOmics: an R package to unravel relationships between two omics datasets. *Bioinformatics* 25, 2855–2856. doi: 10.1093/bioinformatics/btp515
- Long, Q., Rabanal, F. A., Meng, D., Huber, C. D., Farlow, A., Platzer, A., et al. (2013). Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. *Nat. Genet.* 45, 884–890. doi: 10.1038/ng.2678
- Lu, Y., Du, J. I. N., Tang, J., Wang, F., Zhang, J. I. E., Huang, J., et al. (2009). Environmental regulation of floral anthocyanin synthesis in *Ipomoea purpurea*. *Mol. Ecol.* 18, 3857–3871. doi: 10.1111/j.1365-294X.2009.04288.x
- Macel, M., Van Dam, N. M., and Keurentjes, J. J. B. (2010). Metabolomics: the chemistry between ecology and genetics. *Mol. Ecol. Res.* 10, 583–593. doi: 10.1111/j.1755-0998.2010.02854.x
- McQuitty, L. L. (1966). Similarity analysis by reciprocal pairs for discrete and continuous data. *Educ. Psychol. Meas.* 26, 825–831. doi: 10.1177/001316446602600402
- Mitchell-Olds, T., and Schmitt, J. (2006). Genetic mechanisms and evolutionary significance of natural variation in *Arabidopsis*. *Nature* 441, 947–952. doi: 10.1038/nature04878
- Moore, B. D., Andrew, R. L., Külheim, C., and Foley, W. J. (2014). Explaining intraspecific diversity in plant secondary metabolites in an ecological context. *New Phytol.* 201, 733–750. doi: 10.1111/nph.12526
- Morgenthal, K., Wienkoop, S., Scholz, M., Selbig, J., and Weckwerth, W. (2005). Correlative GC-TOF-MS-based metabolite profiling and LC-MS-based protein profiling reveal time-related systemic regulation of metabolite–protein networks and improve pattern recognition for multiple biomarker selection. *Metabolomics* 1, 109–121. doi: 10.1007/s11306-005-4430-9
- Nägele, T. (2014). Linking metabolomics data to underlying metabolic regulation. *Front. Mol. Biosci.* 1:22. doi: 10.3389/fmolb.2014.00022
- Nägele, T., Mair, A., Sun, X., Fragner, L., Teige, M., and Weckwerth, W. (2014). Solving the differential biochemical Jacobian from metabolomics covariance data. *PLoS One* 9:e92299. doi: 10.1371/journal.pone.0092299
- Nägele, T., and Weckwerth, W. (2013). Eigenvalues of Jacobian matrices report on steps of metabolic reprogramming in a complex plant–environment interaction. *Appl. Math.* 4, 44–49. doi: 10.4236/am.2013.48A007
- Oliveros, J. (2007). VENN. An Interactive Tool for Comparing Lists with Venn Diagrams. Available at: <http://bioinfogp.cnb.csic.es/tools/venny/index.html>
- Pigliucci, M. (2010). Genotype-phenotype mapping and the end of the 'genes as blueprint' metaphor. *Philos. Trans. R. Soc. B Biol. Sci.* 365, 557–566. doi: 10.1098/rstb.2009.0241
- Platt, A., Horton, M., Huang, Y. S., Li, Y., Anastasio, A. E., Mulyati, N. W., et al. (2010). The scale of population structure in *Arabidopsis thaliana*. *PLoS Genet.* 6:e1000843. doi: 10.1371/journal.pgen.1000843
- Pracharoenwattana, I., Zhou, W., Keech, O., Francisco, P. B., Udomchalothorn, T., Tschopp, H., et al. (2010). *Arabidopsis* has a cytosolic fumarate required for the massive allocation of photosynthate into fumaric acid and for rapid plant growth on high nitrogen. *Plant J.* 62, 785–795. doi: 10.1111/j.1365-3113X.2010.04189.x
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Sardans, J., Peñuelas, J., and Rivas-Ubach, A. (2011). Ecological metabolomics: overview of current developments and future challenges. *Chemoecology* 21, 191–225. doi: 10.1007/s00049-011-0083-5
- Scherling, C., Roscher, C., Giavalisco, P., Schulze, E. D., and Weckwerth, W. (2010). Metabolomics unravel contrasting effects of biodiversity on the performance of individual plant species. *PLoS One* 5:e12569. doi: 10.1371/journal.pone.0012569

- Shulaev, V., Cortes, D., Miller, G., and Mittler, R. (2008). Metabolomics for plant stress response. *Physiol. Plant.* 132, 199–208. doi: 10.1111/j.1399-3054.2007.01025.x
- Somerville, C., and Koornneef, M. (2002). Timeline - a fortunate choice: the history of *Arabidopsis* as a model plant. *Nat. Rev. Genet.* 3, 883–889. doi: 10.1038/nrg927
- Stekhoven, D. J., and Buhlmann, P. (2012). MissForest–non-parametric missing value imputation for mixed-type data. *Bioinformatics* 28, 112–118. doi: 10.1093/bioinformatics/btr597
- Steuer, R., Kurths, J., Fiehn, O., and Weckwerth, W. (2003). Observing and interpreting correlations in metabolomic networks. *Bioinformatics* 19, 1019–1026. doi: 10.1093/bioinformatics/btg120
- Sun, X., and Weckwerth, W. (2012). COVAIn: a toolbox for uni- and multivariate statistics, time-series and correlation network analysis and inverse estimation of the differential Jacobian from metabolomics covariance data. *Metabolomics* 8, 81–93. doi: 10.1007/s11306-012-0399-3
- Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., et al. (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* 45, D362–D368. doi: 10.1093/nar/gkw937
- Todesco, M., Balasubramanian, S., Hu, T. T., Traw, M. B., Horton, M., Eppe, P., et al. (2010). Natural allelic variation underlying a major fitness trade-off in *Arabidopsis thaliana*. *Nature* 465, 632–636. doi: 10.1038/nature09083
- Turesson, G. (1922). The genotypical response of the plant species to the habitat. *Hereditas* 3, 211–350. doi: 10.1111/j.1601-5223.1922.tb02734.x
- Violle, C., Enquist, B. J., McGill, B. J., Jiang, L., Albert, C. H., Hulshof, C., et al. (2012). The return of the variance: intraspecific variability in community ecology. *Trends Ecol. Evol.* 27, 244–252. doi: 10.1016/j.tree.2011.11.014
- Ward, J. K., and Kelly, J. K. (2004). Scaling up evolutionary responses to elevated CO<sub>2</sub>: lessons from *Arabidopsis*. *Ecol. Lett.* 7, 427–440. doi: 10.1111/j.1461-0248.2004.00589.x
- Weckwerth, W. (2003). Metabolomics in systems biology. *Annu. Rev. Plant Biol.* 54, 669–689. doi: 10.1146/annurev.arplant.54.031902.135014
- Weckwerth, W. (2011a). Green systems biology - From single genomes, proteomes and metabolomes to ecosystems research and biotechnology. *J. Proteomics* 75, 284–305. doi: 10.1016/j.jprot.2011.07.010
- Weckwerth, W. (2011b). Unpredictability of metabolism—the key role of metabolomics science in combination with next-generation genome sequencing. *Anal. Bioanal. Chem.* 400, 1967–1978. doi: 10.1007/s00216-011-4948-9
- Weckwerth, W., Loureiro, M. E., Wenzel, K., and Fiehn, O. (2004a). Differential metabolic networks unravel the effects of silent plant phenotypes. *Proc. Natl. Acad. Sci. U.S.A.* 101, 7809–7814.
- Weckwerth, W., Wenzel, K., and Fiehn, O. (2004b). Process for the integrated extraction, identification and quantification of metabolites, proteins and RNA to reveal their co-regulation in biochemical networks. *Proteomics* 4, 78–83.
- Weigel, D. (2012). Natural variation in *Arabidopsis*: from molecular genetics to ecological genomics. *Plant Physiol.* 158, 2–22. doi: 10.1104/pp.111.189845
- Weigel, D., and Mott, R. (2009). The 1001 genomes project for *Arabidopsis thaliana*. *Genome Biol.* 10:107. doi: 10.1186/gb-2009-10-5-107
- Wienkoop, S., Morgenthal, K., Wolschin, F., Scholz, M., Selbig, J., and Weckwerth, W. (2008). Integration of metabolomic and proteomic phenotypes: analysis of data covariance dissects starch and RFO metabolism from low and high temperature compensation response in *Arabidopsis thaliana*. *Mol. Cell. Proteom.* 7, 1725–1736. doi: 10.1074/mcp.M700273-MCP200
- Wink, M. (2003). Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. *Phytochemistry* 64, 3–19. doi: 10.1016/S0031-9422(03)00300-5
- Winkel-Shirley, B. (2002). Biosynthesis of flavonoids and effects of stress. *Curr. Opin. Plant Biol.* 5, 218–223. doi: 10.1016/S1369-5266(02)00256-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 Nagler, Nägele, Gilli, Fragner, Korte, Platzer, Farlow, Nordborg and Weckwerth. 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 Multi-Omics Analysis Pipeline for the Metabolic Pathway Reconstruction in the Orphan Species *Quercus ilex*

Cristina López-Hidalgo<sup>1†</sup>, Victor M. Guerrero-Sánchez<sup>1†</sup>, Isabel Gómez-Gálvez<sup>1</sup>, Rosa Sánchez-Lucas<sup>1</sup>, María A. Castillejo-Sánchez<sup>2</sup>, Ana M. Maldonado-Alconada<sup>1</sup>, Luis Valledor<sup>3</sup> and Jesus V. Jorrín-Novo<sup>1\*</sup>

<sup>1</sup> Agroforestry and Plant Biochemistry and Proteomics Research Group, Department Biochemistry and Molecular Biology, Universidad de Córdoba, Córdoba, Spain, <sup>2</sup> Instituto de Agricultura Sostenible, Córdoba, Spain, <sup>3</sup> Departamento de Biología de Organismos y Sistemas, Universidad de Oviedo, Oviedo, Spain

## OPEN ACCESS

### Edited by:

Atsushi Fukushima,  
RIKEN, Japan

### Reviewed by:

Raquel Esteban,  
University of the Basque Country  
(UPV/EHU), Spain  
Autar Krishen Mattoo,  
United States Department  
of Agriculture, United States

### \*Correspondence:

Cristina López-Hidalgo  
n12lohic@uco.es  
Jesus V. Jorrín-Novo  
bf1jonoj@uco.es

<sup>†</sup> These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Plant Systems and Synthetic Biology,  
a section of the journal  
Frontiers in Plant Science

**Received:** 02 November 2017

**Accepted:** 11 June 2018

**Published:** 11 July 2018

### Citation:

López-Hidalgo C,  
Guerrero-Sánchez VM,  
Gómez-Gálvez I, Sánchez-Lucas R,  
Castillejo-Sánchez MA,  
Maldonado-Alconada A, Valledor L  
and Jorrín-Novo JV (2018) A  
Multi-Omics Analysis Pipeline for the  
Metabolic Pathway Reconstruction in  
the Orphan Species *Quercus ilex*.  
Front. Plant Sci. 9:935.  
doi: 10.3389/fpls.2018.00935

Holm oak (*Quercus ilex*) is the most important and representative species of the Mediterranean forest and of the Spanish agrosilvo-pastoral “dehesa” ecosystem. Despite its environmental and economic interest, Holm oak is an orphan species whose biology is very little known, especially at the molecular level. In order to increase the knowledge on the chemical composition and metabolism of this tree species, the employment of a holistic and multi-omics approach, in the Systems Biology direction would be necessary. However, for orphan and recalcitrant plant species, specific analytical and bioinformatics tools have to be developed in order to obtain adequate quality and data-density before to coping with the study of its biology. By using a plant sample consisting of a pool generated by mixing equal amounts of homogenized tissue from acorn embryo, leaves, and roots, protocols for transcriptome (NGS-Illumina), proteome (shotgun LC-MS/MS), and metabolome (GC-MS) studies have been optimized. These analyses resulted in the identification of around 62629 transcripts, 2380 protein species, and 62 metabolites. Data are compared with those reported for model plant species, whose genome has been sequenced and is well annotated, including *Arabidopsis*, japonica rice, poplar, and eucalyptus. RNA and protein sequencing favored each other, increasing the number and confidence of the proteins identified and correcting erroneous RNA sequences. The integration of the large amount of data reported using bioinformatics tools allows the Holm oak metabolic network to be partially reconstructed: from the 127 metabolic pathways reported in KEGG pathway database, 123 metabolic pathways can be visualized when using the described methodology. They included: carbohydrate and energy metabolism, amino acid metabolism, lipid metabolism, nucleotide metabolism, and biosynthesis of secondary metabolites. The TCA cycle was the pathway most represented with 5 out of 10 metabolites, 6 out of 8 protein enzymes, and 8 out of 8 enzyme transcripts. On the other hand, gaps, missed pathways, included metabolism of terpenoids and polyketides and lipid metabolism. The multi-omics resource generated in this work will set the basis for ongoing and future studies, bringing the Holm oak closer to model species,

to obtain a better understanding of the molecular mechanisms underlying phenotypes of interest (productive, tolerant to environmental cues, nutraceutical value) and to select elite genotypes to be used in restoration and reforestation programs, especially in a future climate change scenario.

**Keywords:** *Quercus ilex*, omics, metabolome, proteome, transcriptome

## INTRODUCTION

Holm oak (*Quercus ilex*) is the most representative species of the Mediterranean forest, of great importance from an environmental and economic point of view (Rigo and De Caudullo, 2016). Being the key element of the Spanish agro-forestry-pastoral ecosystem “Dehesa,” its fruit, the acorn, is the basis of the staple food of the renowned “black leg” pork (Cantos et al., 2003). *Quercus* spp. have been used in the construction of wine barrels, contributing to the organoleptic properties of the maturing wine (Chira and Teissedre, 2014). The use of acorns in human nutrition and for pharmaceutical purposes has a long history. Employed in ancient civilizations, mainly in Italy and Spain, as food or beverage, nowadays it is far from being consumed like other common nuts (Rakić et al., 2006; Al-Rousan et al., 2013; Meijón et al., 2016). As a nutritionally rich product, and because of its high nutraceutical value, the interest of integrating acorns into the human diet or as a functional food has been raised (Vinha et al., 2016a; Hadidi et al., 2017).

Despite its environmental and economic interest, Holm oak is still an orphan species whose biology is almost unknown, especially at the molecular level. Nevertheless, the work of our group and others, has contributed to acquiring the knowledge on this species, focusing on natural variability (Valero-Galván et al., 2011; Akcan et al., 2017), seed germination and seedling growth (Echevarría-Zomeño et al., 2009; Romero-Rodríguez et al., 2015), physiology (Valero-Galván et al., 2012a), and biotic and abiotic stress-responses (Echevarría-Zomeño et al., 2009; Sghaier-Hammami et al., 2013; Sardans et al., 2014; Simova-Stoilova et al., 2015). The above publications, provide fragmented information, mostly derived from classical biochemical approaches and, to a much lesser extent, those of proteomics (Valero-Galván et al., 2011; Romero-Rodríguez et al., 2014, 2015) transcriptomics (Guerrero-Sanchez et al., 2017), or metabolomics (Rakić et al., 2006; Rabhi et al., 2016; Vinha et al., 2016a; López-Hidalgo, 2017), but lacking a validation and effective integration of the different molecular multilevels.

In spite of their difficulty as orphan, recalcitrant plant species, forest trees, like other experimental plant systems, deserve to be considered at the wide system level, that implicates the use of multidisciplinary approaches, from visual phenotype, to molecular – omics, through physiological and biochemical approaches (Correia et al., 2016; Meijón et al., 2016; Escandón et al., 2017). Systems Biology approaches require the optimization of protocols for both wet and *in silico* analysis.

In this direction, trying to fill this gap with the use of the available high-throughput – omics, its combination and also the implementation of required methodology, we hoped to gain

knowledge on the chemical composition and metabolism of the *Q. ilex* tree species, its variability among and within populations, the effect on endogenous ones and their environmental factors, and the search for molecular markers to select elite genotypes. The lack of information available in public databases on the Holm oak genome, transcriptome (Guerrero-Sanchez et al., 2017), or proteome (Romero-Rodríguez et al., 2014) and the absence of standardized laboratory and analytical protocols make this approach a real challenge.

In this work, we employed a wide range of *in silico* techniques allowing a system biology approach for a non-sequenced species. To obtain the maximum level of biochemical complexity the plant sample employed were multi-organ pools, generated by mixing equal amounts of homogenized tissue from acorn embryo, leaves, and roots. In setting up protocols for transcriptome (NGS-Illumina), proteome (shotgun LC-MS/MS) and metabolome (GC-MS) analysis, and bioinformatic pipelines for annotating transcripts, proteins and metabolites, the Holm oak metabolic pathways were partially reconstructed. This research constitutes the basis for ongoing and future studies to obtain a better understanding of the molecular bases underlying phenotypes of interest (productive, tolerant to environmental cues, nutraceutical value) and the selection of elite genotypes to be used in restoration and reforestation programs, especially in the current climate change scenario. In order to reveal the particularities of the species under study, data have been compared with those reported for model plant species, including *Arabidopsis*, rice, poplar, and eucalyptus.

## MATERIALS AND METHODS

### Plant Material

Mature acorns from Holm oak (*Quercus ilex* L. subsp. *ballota* [Desf.] Samp.) were collected on December 2015 from a tree located in Aldea de Cuenca (province of Córdoba, Andalusia, Spain). Acorns were transported to the lab, sterilized, and germinated as previously reported (Simova-Stoilova et al., 2015). Germinated seeds were sown in pots (500 mL) with perlite and grown in a greenhouse under natural conditions for 4 months up to the 10-leaves stage. Plants were periodically watered at field capacity and once a week with a Hoagland nutrient (Hoagland and Arnon, 1950) solution after the second month. Germinated embryos, cotyledons, leaves, and roots were collected separately, washed with distilled water and frozen in liquid nitrogen. Then, each tissue was separately homogenized in a mortar until a fine powder was obtained and finally stored at  $-80^{\circ}\text{C}$ . The experiments were performed with a pool of fresh weight equivalents of the homogenized tissue from acorn

embryo, cotyledons, leaves, and roots. Depending on the organ, samples from individual trees or plantlets in number of 18 (roots and leaves) to 50 (seed embryos and cotyledons) were collected and mixed. Three independent extractions were performed and only consistent proteins or metabolites, those present in the three replicates, were considered.

## Transcriptomics Analysis

### RNA Extraction and Sequencing

Total RNA was extracted from the frozen homogenized pool tissue following the procedure previously reported by (Guerrero-Sanchez et al., 2017). 50 mg pooled fresh tissue according the procedures previously set up in our laboratory for *Q. ilex* samples was employed (Echevarría-Zomeño et al., 2012). Contaminating genomic DNA was removed by DNase I (Ambion) treatment. Total RNA was quantified spectrophotometrically (DU 228800 Spectrophotometer, Beckman Coulter, TrayCell Hellma GmbH & Co., KG). The high quality and integrity of the RNA preparation were tested electrophoretically (Agilent 2100 Bioanalyzer). Only high-quality RNAs with RIN values >8 and  $A_{260}:A_{280}$  ratios near 2.0 were used for subsequent experiments.

The library construction of cDNA molecules was carried out using Illumina TruSeq Stranded mRNA Library Preparation Kit according to the manufacturer's instructions using 2 µg of total RNA followed by poly-A mRNA enrichment using streptavidin coated magnetic beads and thermal mRNA fragmentation. The cDNA was synthesized, followed by a chemical fragmentation (DNA library) and sequenced in the Illumina HiSeq 2500 platform, using 100 bp paired-end sequencing (De Wit et al., 2012).

### Data Processing

The raw reads obtained from the sequencing platform were preprocessed to retain only high-quality sequences to be subsequently used in the assembly. Each original sequence was quality trimmed considering several parameters (quality trimming based on minimum quality scores, ambiguity trimming to trim off, for example, stretches of Ns, base trim to remove specified number of bases at either 3' or 5' end of the reads). The processed reads were assembled *de novo* using the assembly software MIRA 4.9.6 (Chevreux and Suhai, 1999). Redundancy reduction of the assembled sequenced was carried out by using the CD-HIT 4.6 clustering algorithm (Li et al., 2001, 2002).

### Gene Ontology

Assembled sequences were blasted against UniRef90 (UniProt<sup>1</sup>) using the software Sma3s (Casimiro-Soriguer et al., 2017) in order to obtain the annotated sequences with the most probable gene name and protein description, EC numbers for enzymes, GO terms, and UniProt keywords and pathways. In addition, their functions were identified using MERCATOR<sup>2</sup>.

<sup>1</sup> <http://ftp.uniprot.org/pub/databases/uniprot/uniref/uniref90/>

<sup>2</sup> <http://www.plabipd.de/portal/mercator-sequence-annotation/>

## Proteomics Analysis

### Protein Extraction and Digestion

Proteins were extracted from the frozen homogenized pool tissue by using the TCA-acetone-phenol protocol as reported in Jorrin-Novo et al. (2014). Protein extracts [600–1000 ng BSA equivalents quantified with Bradford assay (Bradford, 1976)] were subjected to Orbitrap analysis after SDS-PAGE (12%) pre-fractionation. Electrophoresis was stopped when the sample entered the resolving gel, so that a unique protein band was revealed after Coomassie staining (Pascual et al., 2017).

Protein bands were manually excised, destained, and digested with trypsin Sequencing grade (Roche) as is described in Castillejo et al. (2015) with minor modifications. Briefly, gel plugs were destained by incubation (twice for 30 min) with a solution containing 100 mM ammonium bicarbonate (AmBic)/50% acetonitrile (AcN) at 37°C. Then, they were dehydrated with AcN and incubated in 100 mM AmBic containing first 20 mM DTT for 30 min, and then in the same solution containing 55 mM Iodoacetamide instead DTT for 30 min. They were washed with 25 mM AmBic and 25 mM AmBic/50% AcN two times each. After dehydration in AcN, the trypsin at a concentration of 12.5 ng/µl was added in a buffer containing 25 mM  $\text{NH}_4\text{HCO}_3$ , 10% AcN and 5 mM  $\text{CaCl}_2$ , and the digestion proceeded at 37°C for 12 h. Digestion was stopped, and peptides were extracted from gel plugs by adding 10 µL of 1% (v/v) trifluoroacetic acid (TFA) and incubating for 15 min.

### Shotgun LC-MS Analysis

Nano-LC was performed in a Dionex Ultimate 3000 nano UPLC (Thermo Scientific) with a C18 75 µm × 50 Acclaim Pepmap column (Thermo Scientific). The peptide mix was previously loaded on a 300 µm × 5 mm Acclaim Pepmap precolumn (Thermo Scientific) in 2% AcN/0.05% TFA for 5 min at 5 µL/min. Peptide separation was performed at 40°C for all runs. Mobile phase buffer A was composed of water, 0.1% formic acid. Mobile phase B was composed of 80% AcN, 0.1% formic acid. Samples were separated during a 60-min gradient ranging from 96% solvent A to 90% solvent B and a flow rate of 300 nL/min.

Eluted peptides were converted into gas-phase ions by nano electrospray ionization and analyzed on a Thermo Orbitrap Fusion (Q-OT-qIT, Thermo Scientific) mass spectrometer operated in positive mode. Survey scans of peptide precursors from 400 to 1500 *m/z* were performed at 120K resolution (at 200 *m/z*) with a  $4 \times 10^5$  ion count target. Tandem MS was performed by isolation at 1.2 Da with the quadrupole, CID fragmentation with normalized collision energy of 35, and rapid scan MS analysis in the ion trap. The AGC ion count target was set to  $2 \times 10^3$  and the maximum injection time was 300 ms. Only those precursors with charge state 2–5 were sampled for MS<sup>2</sup>. The dynamic exclusion duration was set to 15 s with a 10 ppm tolerance around the selected precursor and its isotopes. Monoisotopic precursor selection was turned on. The instrument was run in top 30 mode with 3 s cycles, meaning that the instrument would continuously perform MS<sup>2</sup> events until a maximum of top 30 non-excluded precursors or 3 s, whichever was shorter.



## Protein Identification

Spectra were processed using the SEQUEST algorithm available in Proteome Discoverer® 1.4 (Thermo Scientific, United States). The following settings (Romero-Rodríguez et al., 2014) were used: precursor mass tolerance was set to 10 ppm and fragment ion mass tolerance to 0.8 Da. Only charge states + 2 or greater were used. Identification confidence was set to a 5% FDR and the variable modifications were set to: oxidation of methionine and the fixed modifications were set to carbamidomethyl cysteine formation. A maximum of two missed cleavages were set for all searches. The protein identification, was carried out against the annotated *Q. ilex* transcriptome, previously described. A six-frame translation for each sequence in the transcriptome was performed by using EMBOSS (Rice et al., 2000), filtering and keeping peptides longer than 50 amino acids. Considering the identified proteins, the protein peak areas were normalized and missing values corrected. Mean values and standard deviation (SD), as well as the coefficient of variation (CV) of the peak areas of protein species were determined for three independent analysis (Supplementary Table S8). The remaining sequences were used as a database for the protein identifications and their functions were identified using MERCATOR (Lohse et al., 2014).

## Metabolomics Analysis

### Metabolite Extraction

Metabolites were extracted from plant tissue as described by Valledor et al. (2014), with three independent extractions. A buffer containing 600 µL of cold methanol: chloroform: water (5:2:2) was added to 15 mg of frozen tissue, vortexed (10 s), and the mixture sonicated (ultrasonic bath, 40 KHz for 10 min). After centrifugation (4°C, 4 min, 20,000 × g) the supernatant was transferred to new tubes containing 400 µL of cold chloroform: water (1:1). For phase separation, the tubes were centrifuged (4°C, 4 min, 20,000 × g). The upper (polar) and the lower (apolar) phases were re-extracted with 200 µL of cold chloroform (upper) and water (lower), respectively. After combining on one hand the water: methanol (upper) and, on the other the chloroform (lower) phases, they were vacuum dried at 25°C (Speedvac, Eppendorf Vacuum Concentrator Plus/5301).

### GC-MS Analysis

GC-MS analysis was performed as reported Furuhashi et al. (2012) and Meijón et al. (2016) with some modifications. Polar (water: methanol dissolved) metabolites were derivatized by re-suspending the dried extract in 20 µL of anhydrous pyridine containing 40 mg/mL of methoxyamine hydrochloride. The mixture was incubated at 30°C for 30 min under agitation. Next, 60 µL of *N*-methyl-*N*-trimethylsilyl trifluoroacetamide (MSTFA) was added, samples incubated at 60°C for 30 min, centrifuged (3 min, 20,000 × g), and cooled to room temperature. Then, 80 µL of the supernatant was transferred to GC-microvials. Apolar (chloroform solubilized) metabolites were methylesterified with 295 µL tert-methyl-Butyl-Ether (MTBE), and 5 µL of trimethylsulfonium hydroxide solution (TMSH) for 30 min at room temperature. The tubes were centrifuged (3 min, 20,000 × g) to remove insoluble particles before transferring the supernatants to GC-microvials.

Polar metabolites were resolved and analyzed with a Gas Chromatograph/Mass Spectrometer Agilent 5975B GC/MSD. Inlet temperature was set at 230°C. Samples were injected in discrete randomized blocks with a 1.2 mL/min flow rate. GC separation was performed splitless on a HP-5MS capillary column (30 m × 0.25 mm × 0.25 mm) (Agilent 19091J-433) over a 70–76°C gradient at 0.75°C/min, 76–180°C gradient at 6°C/min, 180–200°C gradient at 3.5°C/min, and then to 310°C at 6°C/min. The mass spectrometer operated in electron-impact (EI) mode at 70 eV in a scan range of *m/z* 40–800. For apolar metabolites a different temperature gradient was employed: 80–190°C at 8°C/min, 190–220°C at 5°C/min, and then to 270°C at 5°C/min. The mass spectrometer was operated in EI mode at 70 eV in a scan range of *m/z* 40–600.

### Metabolite Identification

Metabolites were “tentatively assigned” based on GC retention times (RT) and *m/z* values (Supplementary Tables S1, S2) through searches in different databases, including the Gölm Metabolome Database (Nielsen and Jewett, 2007), Alkane, Fiehn library 1 y 2 (Kind et al., 2010), GC-TSQ, MoSys, and NIST/EPA/NIH Mass Spectral Library. Three different softwares were used for metabolite identification: MZmine 2 (2.24 version<sup>3</sup>) (Pluskal et al., 2010), AMDIS software (2.66 version<sup>4</sup>), and NIST.MS Search (2.01 version<sup>5</sup>). Mean values and SD, as well as the CV of the peak areas of metabolites were determined for three independent extraction (Supplementary Table S2). Moreover, the metabolites were annotated using the KEGG compound reference database<sup>6</sup>. Metabolomics pathways of each metabolite (Supplementary Table S3) were searched against KEGG pathway maps<sup>7</sup>. For other general biological networks, we employed MapMan (3.5.1 version<sup>8</sup>).

## Interspecies Comparison

The annotated *Q. ilex* transcriptome was compared against the complete *in silico* proteomes of *Arabidopsis thaliana* (UP000006548<sup>9</sup>), *Oryza sativa subsp. japonica* (UP000059680<sup>10</sup>), *Populus trichocarpa* (UP000006729<sup>11</sup>), and *Eucalyptus grandis* (UP000030711<sup>12</sup>) in order to elucidate the unique and shared sequences. This comparison was performed by using BLAST<sup>13</sup> with blastX alignment with an *e*-value of 10<sup>−10</sup>. Also, the EC numbers of each proteome were contrasted to achieve a complete picture of the metabolic pathways coverage differences among proteomes studied in previously mentioned species (Supplementary Table S4). For the comparison, we represented a

<sup>3</sup><http://mzmine.github.io/>

<sup>4</sup><http://www.amdis.net/>

<sup>5</sup><http://chemdata.nist.gov/mass-spc/ms-search/>

<sup>6</sup><http://www.genome.jp/kegg/compound/>

<sup>7</sup>[https://www.genome.jp/kegg/tool/map\\_pathway1.html](https://www.genome.jp/kegg/tool/map_pathway1.html)

<sup>8</sup><http://mapman.gabipd.org/web/guest/mapman/>

<sup>9</sup><http://www.uniprot.org/proteomes/UP000006548>

<sup>10</sup><http://www.uniprot.org/proteomes/UP000059680>

<sup>11</sup><http://www.uniprot.org/proteomes/UP000006729>

<sup>12</sup><http://www.uniprot.org/proteomes/UP000030711>

<sup>13</sup><https://blast.ncbi.nlm.nih.gov/Blast.cgi>

Venn diagram plotted using VennDiagram R package (Chen and Boutros, 2011).

## Integrated Pathway

By using MERCATOR web application<sup>14</sup> (Supplementary Tables S5, S6) (Lohse et al., 2014), we could assign MapMan “Bins” to arbitrary transcript or protein input sequences (Usadel et al., 2009). The output was a text file mapping each input (proteins or transcripts) identifier to one or more Bins by searching a variety of reference databases (TAIR Release 10, SwissProt/UniProt Plant Proteins, Clusters of Orthologous Eukaryotic Genes Database (KOG), Conserved Domain Database (CDD), and InterProScan). The functional predictions generated could directly be used as a “mapping file” for the high-throughput data visualization and meta-analysis software MapMan (3.5.1 version<sup>15</sup>). The ImageAnnotator module allowed us to visualize the data on a gene-by-gene basis on schematic diagrams (maps) of the biological processes described.

## RESULTS AND DISCUSSION

This paper reports the study and view of the metabolism as it occurs in Holm oak, the most representative and valuable forest tree species in the Mediterranean region. For that purpose, a biological sample containing equal fresh weight amount of the different organs as starting plant material and a combination of high-throughput, -omics approaches (transcriptomics, proteomics, and metabolomics) as analytical tools were used. As each analytical platform has its own limitations (Schrimpe-Rutledge et al., 2016; Tian et al., 2016; Viant et al., 2017), is their integration that will provided more confident biological knowledge of them.

The Systems Biology approach for research with species that, like Holm oak are orphan and recalcitrant is very challenging (Abril et al., 2011), and it required the optimization of experimental protocols and, more limitative, the creation of custom-made databases, and pipelines. Beyond the reconstruction of different metabolic pathways as they may occur in Holm oak, and the comparison with model plant species (*A. thaliana*, *O. sativa* subsp. *japonica*, *P. trichocarpa*, and *E. grandis*) we aimed to prove that employing state-of-the-art instrumentation and a similar workflow to those employed in model species is feasible, even though quite uncommon in the current literature.

## Transcriptome Analysis

The first transcriptome of *Q. ilex* has recently been reported. For that reason, the Illumina Hiseq 2500 platform was employed to analyze the tissue mix sample, resulting in 119889 contigs, and 31973 Blast2GO annotated transcripts (Guerrero-Sanchez et al., 2017). The number of annotated sequences have been increased to 62628 after a UniRef90

database search through Sma3s software (Muñoz-Mérida et al., 2014; Casimiro-Soriguer et al., 2017). Among them, 27089 sequences corresponded to unique genes. Comparatively, Sma3s performed faster than Blast2GO and allowed more elaborated results, including functional categories, such as biological processes, cellular components or molecular functions (Supplementary Figures S1–S3). The total transcriptome sequences were categorized in 35 MERCATOR functional plant categories. The result of this categorization showed a high percentage (41.8%) of non-assigned transcripts (Figure 1). Response to stress and biosynthetic process, and the nucleus and plastids, were, respectively, the biological processes and organelles most represented (Supplementary Figures S1, S2). With respect to molecular functions, ion binding and kinase activity were those most abundant, with around 11225 and 6372 sequences, respectively (Supplementary Figure S3).

The number of annotated transcripts, 62628, is double that previously found for the close relative *Q. robur* (38292 sequences; Tarkka et al., 2013), similar to the figure of 27655 protein-coding genes in *Arabidopsis* (35386 identified proteins; Araport11<sup>16</sup>), and below the 82190 unique transcripts corresponding to 34212 genes also reported in *Arabidopsis* by Zhang et al. (2017).

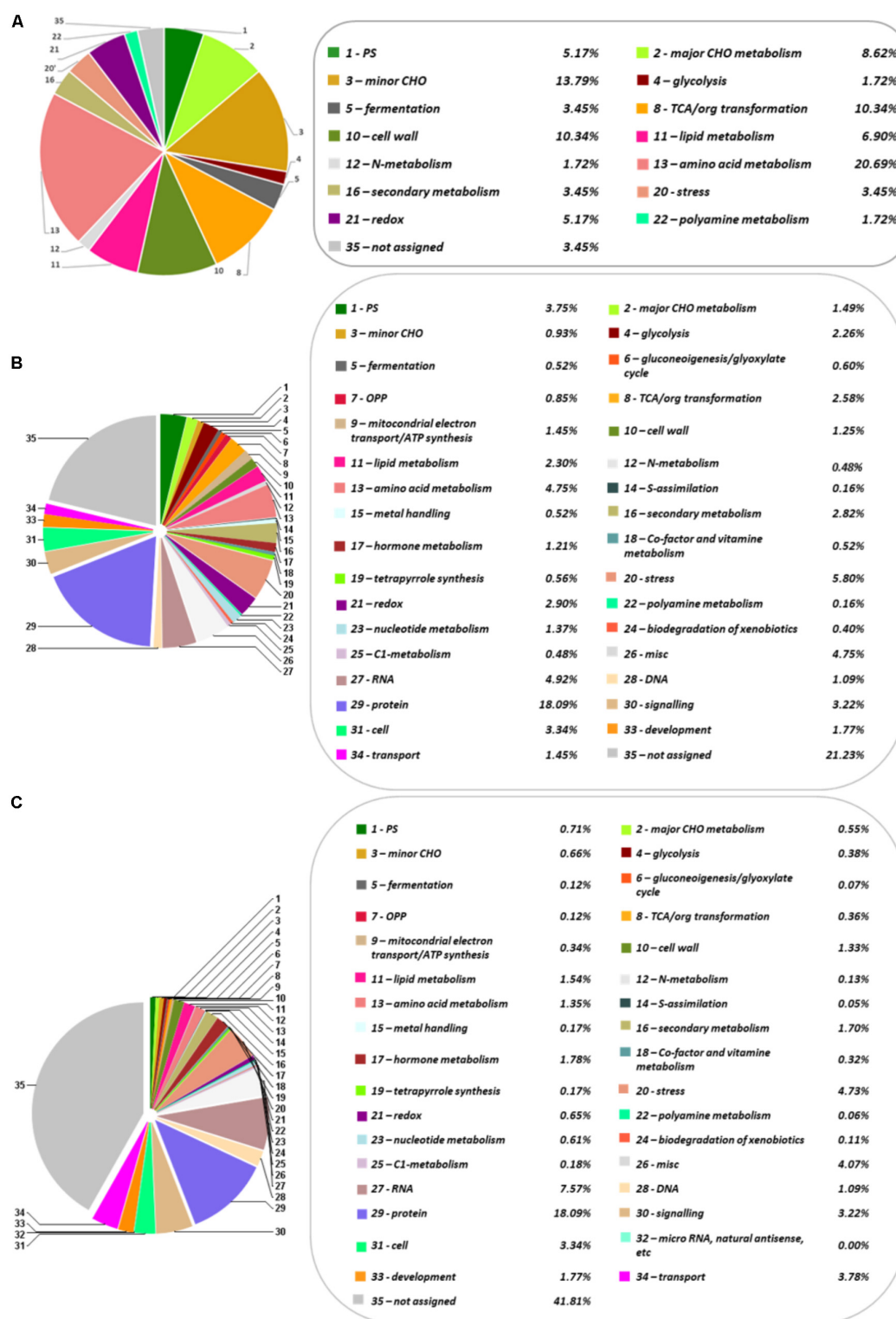
The annotated sequences in *Q. ilex* transcriptome were compared with the *in silico* proteomes of *A. thaliana*, *O. sativa* subsp. *japonica*, *P. trichocarpa*, and *E. grandis* (UniProt) to elucidate the unique and shared sequences. The comparative results are shown in Supplementary Table S2. The highest percentage of similarity corresponded to *P. trichocarpa* (91.7%), and lowest to *O. sativa* subsp. *japonica* (77.8%), with intermediate values for *E. grandis* (88.5%) and *A. thaliana* (85.6%). The percentage of similarity correlated with the phylogenetic distances among the compared species as reported by The Angiosperm Phylogeny Group III (2009) (Figure 2).

Among the annotated transcripts, 2103 corresponded to enzyme transcript products. These enzymes were assigned to 123 KEGG metabolic pathways (Supplementary Table S3). The most represented pathways (Table 2) were: the carbohydrate metabolism (starch and sucrose metabolism and glycolysis/gluconeogenesis, with 26 and 30 enzyme transcripts, respectively). Also, the amino acids metabolism, primarily the cysteine and methionine metabolism, where 37 enzyme transcripts were detected. This pathway has an important role in plants. Cysteine constitutes the sulfur donor for the biosynthesis of methionine, phytochelatin, sulphydryl compounds, glutathione, and coenzymes. The homeostasis of sulfur metabolism in trees is more robust than in herbaceous plants. Also, a greater change in conditions to initiate a response in trees is required (Rennenberg et al., 2007). This fact is coherent with the requirement for highly flexible defense strategies in woody plant species because of longevity. In addition, the lipid metabolism (glycerophospholipid metabolism with 32 enzyme transcripts) has an important function as a mediator in hormone signal transduction in plants (Janda et al., 2013).

<sup>14</sup><http://www.plabipd.de/portal/mercator-sequence-annotation/>

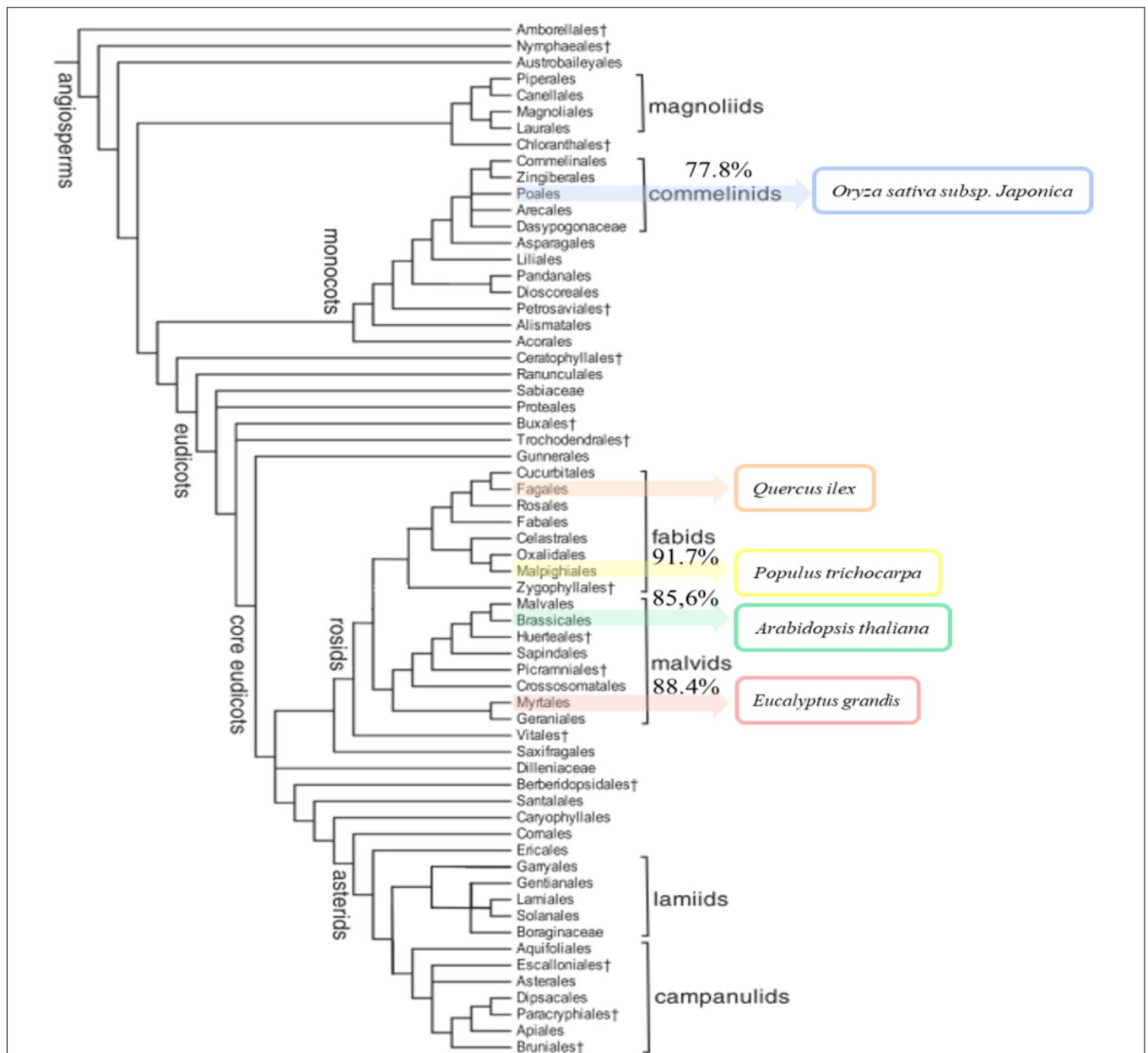
<sup>15</sup><https://mapman.gabipd.org/mapman-download/>

<sup>16</sup><https://www.arabidopsis.org/>



**FIGURE 1 |** Functional categorization and distribution in percentage of the identified metabolites, proteins and transcripts, according to the categories established by MERCATOR. **(A)** Metabolome. **(B)** Transcriptome. **(C)** Proteome. The pie charts show different functional categories: PS (Photosynthesis), major CHO metabolism, minor CHO metabolism, glycolysis, fermentation, gluconeogenesis/glyoxylate cycle, OPP (Oxidative Pentose Phosphate), TCA/org transformation, mitochondrial electron transport/ATP synthesis, cell wall, lipid metabolism, N-metabolism, amino acid metabolism, S-assimilation, metal handling, secondary metabolism, hormone metabolism, co-factor and vitamin metabolism, tetrapyrrole synthesis, stress, redox, polyamine metabolism, nucleotide metabolism, biodegradation of xenobiotics, C1-metabolism, miscellanea, RNA, DNA, protein, signaling, cell, micro RNA, natural antisense, etc., development, transport, and not assigned.





**FIGURE 2 |** Phylogenetic tree of angiosperms. The tree shows the five-species compared (*Arabidopsis thaliana*, *Eucalyptus grandis*, *Oryza sativa subsp. japonica*, *Populus trichocarpa*, and *Quercus ilex*). The sequence similarity of species coincides with the classification in the phylogenetic tree. Species are ranked from highest to lowest similar to *Q. ilex*: *P. trichocarpa* (91.7%), *E. grandis* (88.4%), *A. thaliana* (85.6%), and *O. sativa subsp. japonica* (77.8%).

## Proteome Analysis

The protein profile of the *Q. ilex* tissue mix sample was analyzed using a shotgun proteomics platform. Protein extracts were obtained by using a TCA-acetone-phenol protocol. After trypsin digestion, peptides were subjected to UPLC-Q-OT-qIT MS. The resulting peptides and corresponding proteins were identified by matching MS and MS/MS *m/z* data against the protein database resulting from the six-frame translation of the *Q. ilex* transcriptome. The employment of species specific databases instead of generic Viridiplantae ones

improved the number and confidence of the identifications, as previously published (Romero-Rodríguez et al., 2014). By using Viridiplantae (SwissProt), 891 proteins were identified. Nevertheless, with our custom-built specific database, 58584 peptides were detected corresponding to 2830 proteins (with at least one unique peptide (**Supplementary Tables S7, S8**). Mean, SD, and CV (%) values of normalized identified protein peak areas were determined for three replicates (**Supplementary Table S8**). The mean of the CV obtained was 36.75% (**Supplementary Table S8**), which was slightly higher than the CV mean previously

described using a 2-DE gel analysis (28.9%) (Jorge et al., 2005, 2006). This is due to the number proteins, considering that this number is much lower in a 2-DE gel analysis and usually highly represented than in a shotgun LC-MS/MS. However, despite having a slightly higher value of CV, the shotgun LC-MS/MS shows greater sensitivity and wide dynamic range. Proteins were categorized in 34 MERCATOR functional plant categories (**Figure 1**). 21.2% of the proteins was not assigned to a functional plant category. Up-to 18.1% proteins were related to protein fate (assembly, folding, degradation, and protein posttranslational modifications), this group being the one most represented.

The Holm oak proteome was filtered manually looking for proteins corresponding to enzymes based on the EC number. This resulted in 228 enzyme proteins, corresponding to 10% of the protein species with EC deduced from the *in silico* predicted Holm oak transcriptome (2103 enzyme proteins) and around 20–50% of the enzymes predicted for the sequenced *A. thaliana* and *O. sativa subsp. japonica* systems at UniProt.

The proteins identified were assigned to 93 KEGG metabolic pathways (**Supplementary Table S3**). The most represented pathways were: the carbohydrate metabolism (starch and sucrose metabolism and glycolysis/gluconeogenesis) and the amino acids metabolism (**Table 2**). The least represented one was the enzymes related to transcription (**Supplementary Table S3**). These figures are much higher than those previously reported for *Q. ilex* and other forest tree species (Valero-Galván et al., 2012b; Pascual et al., 2017; Szuba and Lorenc-Plucińska, 2017), maybe due to the use of the powerful LTQ-Orbitrap mass instrument (Kalli et al., 2013) and the search in custom-built specific database.

Out of the 228 enzyme proteins identified, 23 were specific for Holm oak, and 202, 157, 88, and 87, shared with, respectively, *A. thaliana*, *O. sativa subsp. japonica*, *P. trichocarpa*, and *E. grandis* (**Figure 2**). 84 enzymes were common to all the species, and 471, and 35 specific for *A. thaliana* and *O. sativa subsp. japonica*. It is worthnoting that, for *P. trichocarpa* and *E. grandis* no unique enzymes were found, this proving the

quality and validity of our data, with, consequently, a more complete annotated transcriptome and proteome database. Holm oak unique enzymes were related to the biosynthesis of hormones and secondary metabolites. They included those involved in the zeatin biosynthetic pathway (*ath00908*), such as *cis*-zeatin O-beta-D-glucosyltransferase (EC:2.4.1.215) and zeatin O-beta-D-xylosyltransferase (EC:2.4.2.40). Zeatin, one of the growth promoting hormones, is the predominant xylem-mobile cytokinin in many plant species (Kamboj et al., 1999). In the Holm oak unique enzymes involved in the secondary metabolism [6'-deoxychalcone synthase (EC:2.3.1.170) and prenylcysteine oxidase (EC:1.8.3.5)] were involved in flavonoid biosynthesis and terpenoid backbone biosynthesis, respectively. This is not surprising as secondary metabolites are species specific. Thus, in Holm oak, the flavonoids epicatechin gallate and epigallocatechin were found (Vinha et al., 2016b).

The 84 enzyme proteins common to the five-species corresponded mostly to pathways of the central metabolism, such as those of starch and sucrose (e.g., sucrose synthase, EC: 2.4.1.13, and glucose-6-phosphate isomerase, EC: 5.3.1.9), glycolysis and gluconeogenesis [e.g., phosphoglycerate kinase (EC:2.7.2.3) and pyruvate kinase (EC:2.7.1.40)], and citrate cycle [e.g., malate dehydrogenase (EC:1.1.1.37), pyruvate dehydrogenase (EC:1.2.4.1), and aconitate hydratase (EC:4.2.1.3)].

The 228 enzyme proteins identified belonged to 109 pathways, with some of them being represented by only one enzyme [e.g., caffeine metabolism (*ath00232*) and arachidonic acid metabolism (*ath00590*)] and up to 20 enzymes [e.g., carbon fixation in photosynthetic organisms (*ath00710*)]. Analysis of the UniProt *in silico* enzyme proteome revealed 106 and 107 pathways for, respectively, *P. trichocarpa* and *E. grandis*, with the figure being higher for *A. thaliana* (121 pathways) and *O. sativa subsp. japonica* (112 pathways) (**Supplementary Table S8**).

The pathways most represented in Holm oak were those of the intermediate and central metabolism, including

**TABLE 1 |** Metabolite families from GC-MS data of *Quercus ilex*.

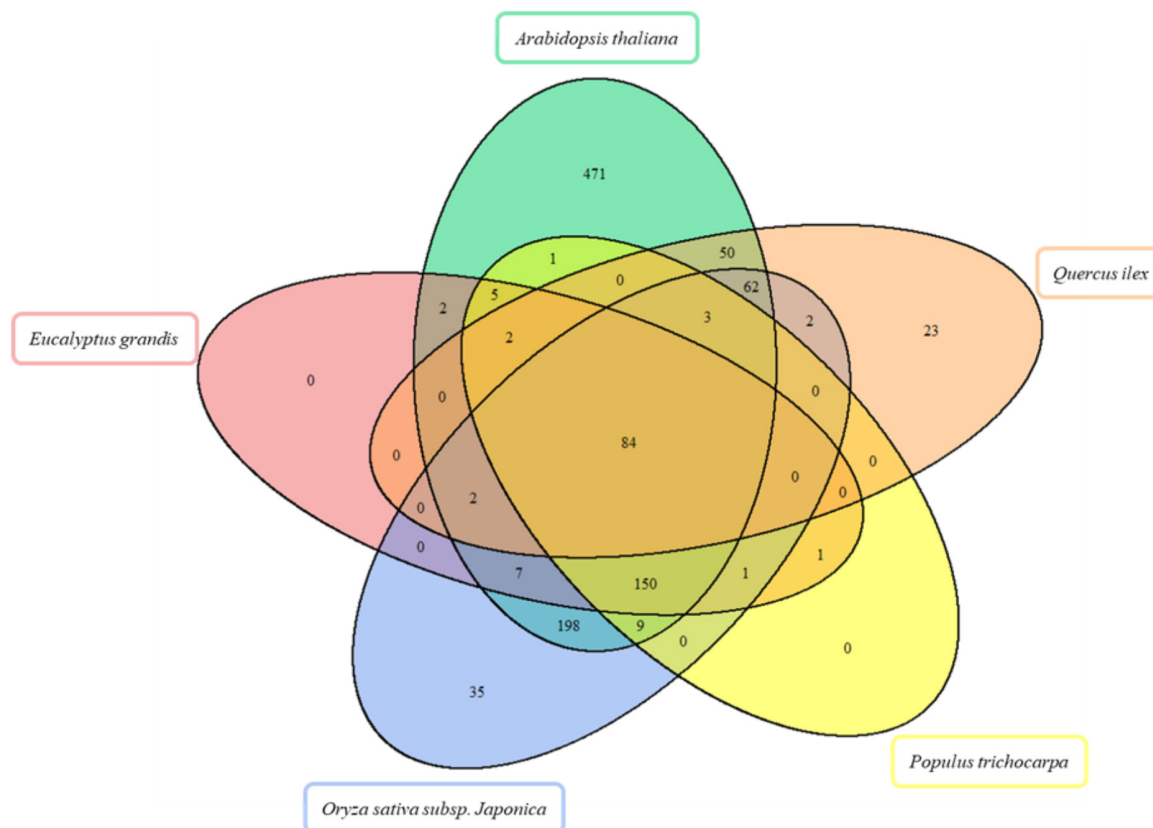
Nature of the compounds	Metabolite name
Amino acids	L-Glutamate (C00025), L-aspartate (C00049), L-alanine (C00041), L-asparagine (C00152), L-serine (C00065), L-threonine (C00188), L-leucine (C00123), L-valine (C00183), L-isoleucine (C00407), L-proline (C00148), L-phenylalanine (C00079)
Organic acids	Ascorbate (C00072), pyruvate (C00022), L-lactate (C00186), succinate (C00042), fumarate (C00122), malate (C00149), citrate (C00158), aconitate (C00417), gluconolactone (C00198), D-glycerate (C00258), glucarate (C00818), galactarate (C00879), maleate (C01384), salicylate (C00805), pyroglutamic acid (C01879), oxalate (C00209), gallate (C00627), quinate (C00296), D-ribonate (C01685)
Carbohydrates	D-Glucose (C00031), L-arabinose (C00259), D-xylulose (C00310), D-galacturonate (C00333), D-fructose (C00095), L-sorbose (C00247), mannitol (C00392), L-rhamnose (C00507), D-sorbitol (C00794), sucrose (C00089), D-galactose (C00124), melibiose (C05402), myo-inositol (C00137), D-glucose 6-phosphate (C00092), maltose (C00208), maltotriose (C01835), D-cellobiose (C00185), D-galactonate (C00880), D-erythrose (C01796)
Polyols	Glycerol (C00116), viburnitol (C08259)
Fatty acids	Palmitic acid (C00249), oleic acid (C00712), stearic acid (C01530), linoleic acid (C01595)
Phenolic compounds (flavonoids)	Catechin (C06562), epigallocatechin (C12136)
Others	Urea (C00086), 4-aminobutanoate (GABA) (C00334), tridecane (C13834), anthraquinone (C16207)

Six main chemical families of metabolites are represented. Carbohydrates (19), organic acids (19), amino acids (11), fatty acids (4), polyols (2), phenolic compounds (2), and four unique compound classes (others). Data in the brackets are KEGG compound identifier of each metabolite.

TABLE 2 | Number of metabolites and enzymes (proteomic and transcriptomic level) in KEGG pathways.

Pathways	Metabolites	Proteins	Transcripts
Carbohydrate metabolism			
Glycolysis/gluconeogenesis ( <i>ath00010</i> )	Pyruvate, D-glucose, L-lactate	3	20
Glyoxylate and dicarboxylate metabolism ( <i>ath00630</i> )	Pyruvate, L-glutamate, succinate, L-serine, malate, citrate, glycolate, oxalate, glycerate, aconitate	10	16
Citrate cycle (TCA cycle) ( <i>ath00020</i> )	Pyruvate, succinate, fumarate, malate, citrate, aconitate	6	9
Amino sugar and nucleotide sugar metabolism ( <i>ath00520</i> )	D-Glucose, L-arabinose, D-galacturonate	3	12
Starch and sucrose metabolism ( <i>ath00500</i> )	D-Glucose, sucrose, D-glucose 6-phosphate, D-fructose, cellobiose, maltose	6	18
Pentose phosphate pathway ( <i>ath00030</i> )	Pyruvate, D-glucose, gluconolactone, glycerate	4	7
Galactose metabolism ( <i>ath00052</i> )	D-Glucose, sucrose, D-fructose, glycerol, D-galactose, myo-inositol, D-sorbitol, D-galactonate, melibiose	9	9
Amino acid metabolism			
Alanine, aspartate, and glutamate metabolism ( <i>ath00250</i> )	Pyruvate, L-glutamate, L-alanine, succinate, L-aspartate, fumarate, L-asparagine, citrate, 4-aminobutanoate (GABA)	9	9
Cysteine and methionine metabolism ( <i>ath00270</i> )	Pyruvate, L-alanine, L-aspartate, L-serine	4	10
Glycine, serine, and threonine metabolism ( <i>ath00260</i> )	Pyruvate, L-aspartate, L-serine, L-threonine, glycerate	5	11
Phenylalanine metabolism ( <i>ath00360</i> )	Pyruvate, succinate, L-phenylalanine, fumarate, salicylate	5	4
Lipid metabolism			
Biosynthesis of unsaturated fatty acids ( <i>ath01040</i> )	Palmitic acid, oleic acid, stearic acid, linoleic acid	4	3
Energy metabolism			
Carbon fixation in photosynthetic organisms ( <i>ath00710</i> )	Pyruvate, L-alanine, L-aspartate, malate	4	20
Biosynthesis of other secondary metabolites			
Phenylpropanoid biosynthesis ( <i>ath00940</i> )	L-Phenylalanine, gallate	2	7

Pathways according to the KEGG pathway maps based on *Arabidopsis thaliana*. The *Arabidopsis* pathway identifiers are in brackets. The table shows the most representative pathways. The complete list of pathways is in the **Supplementary Material Table S9**.



**FIGURE 3 |** Venn diagram for the comparison of enzymes in *Arabidopsis thaliana*, *Eucalyptus grandis*, *Oryza sativa subsp. japonica*, *Populus trichocarpa* *in silico* proteomes, and *Quercus ilex* proteome. The Venn diagram shows the overlap of enzymes detected.

glyoxylate and dicarboxylate metabolism (*ath00630*) with 16 enzyme proteins and amino sugar and nucleotide sugar metabolism (*ath00520*) with 12 enzyme proteins (**Table 2**). For the glycolysis (**Supplementary Figure S4**), just as an example, there were only two enzyme proteins non-detected: phosphofructokinase (EC:2.7.1.11) and phosphoglycerate mutase (EC:5.4.2.12) (**Supplementary Table S9**). These results are more complete than the ones found from the *in silico* analysis of the other two woody plants used for comparisons *P. trichocarpa* and *E. grandis*, with only 5 out of the 10 glycolytic enzymes.

## Metabolome Analysis

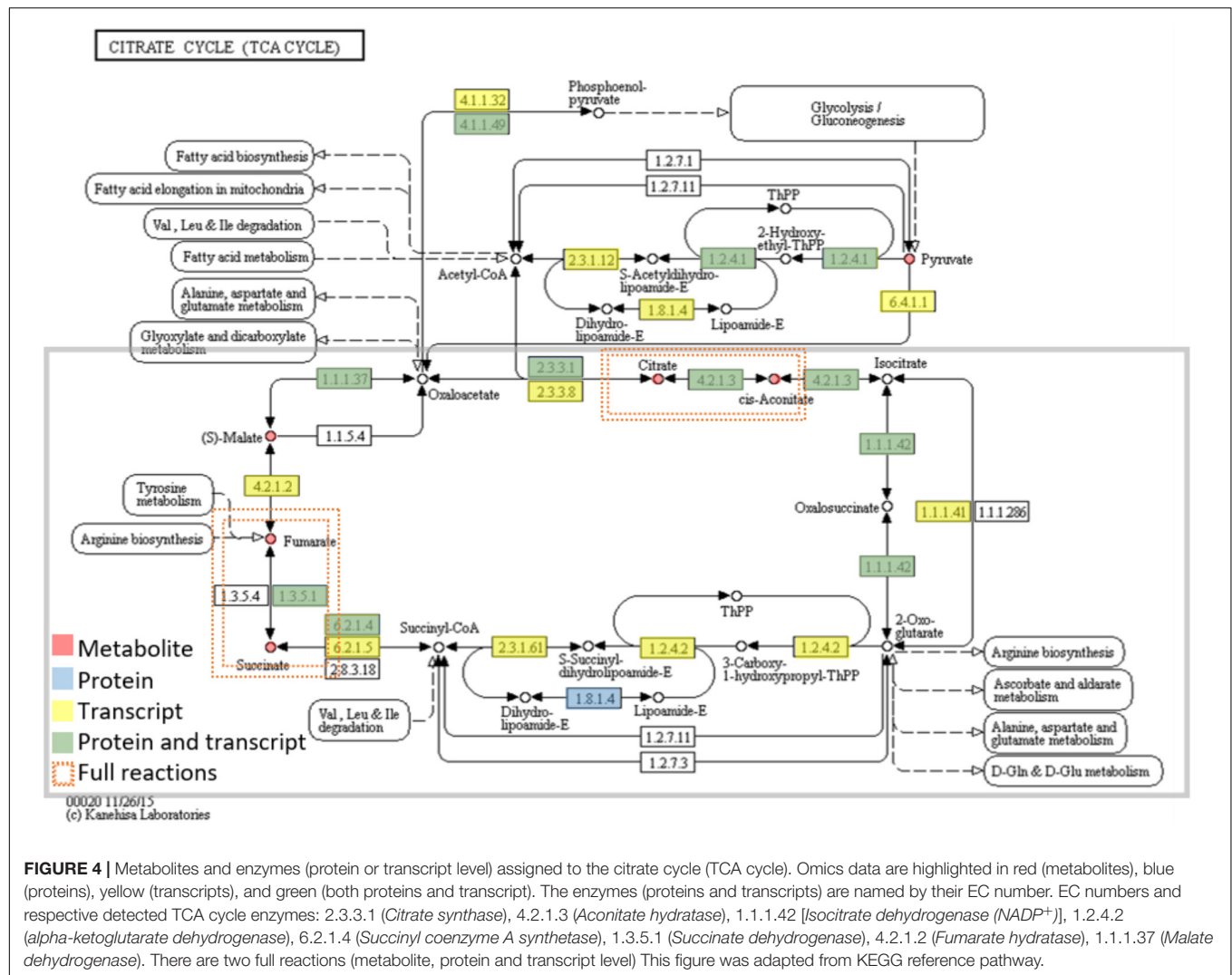
The metabolites present in the pooled samples were analyzed by using GC-q-MS. Two different extraction solvents, methanol:water and chloroform, were, respectively, used for compounds of different polarities. Up to 155 and 19 peaks were resolved by gas chromatography using the above mentioned solvents. A complete list of the identified compounds with their respective RT and the mass-to-charge ratios (*m/z*) is included in **Supplementary Tables S1, S2**. From the *m/z* values, and after a search in seven public databases (Alkane, Fiehn library 1 and 2, Gölm Metabolome Database, GC-TSQ, MoSys, and NIST/EPA/NIH Mass

Spectral Library) a total of 62 compounds were identified, 57 in the methanol:water extract and 5 in the chloroform one. The normalized peak areas of the metabolites were employed for the mean, SD, and CV determinations. The average of the CV obtained (13.70%) was lower than the obtained with proteins data (36.75%), revealing the existence of a greater variability in proteins analysis. The higher CV could be related with the higher number and diversity of identified proteins versus the metabolites identified.

Identified compounds were in the 60–500 Da and mostly belonged to the primary metabolism (59), with only three being secondary metabolites (catechin, epigallocatechin, and anthraquinone). The identified metabolites were grouped in six chemical families according to the KEGG database<sup>17</sup>, including carbohydrates (19), organic acids (19), amino acids (11), fatty acids (4), polyols (2), and phenolic compounds (2) (**Table 1**). The family most represented was that of organic acids (19) and carbohydrates (19), followed by amino acids (11). Fatty acids (4) and phenolic compounds (2) were much less represented. They were included in at least 64 different KEGG pathways (**Supplementary Table S3**), and in 15

<sup>17</sup><http://www.genome.jp/kegg/compound/>





**FIGURE 4 |** Metabolites and enzymes (protein or transcript level) assigned to the citrate cycle (TCA cycle). Omics data are highlighted in red (metabolites), blue (proteins), yellow (transcripts), and green (both proteins and transcript). The enzymes (proteins and transcripts) are named by their EC number. EC numbers and respective detected TCA cycle enzymes: 2.3.3.1 (*Citrate synthase*), 4.2.1.3 (*Aconitate hydratase*), 1.1.1.42 [*isocitrate dehydrogenase (NADP<sup>+</sup>)*], 1.2.4.2 (*alpha-ketoglutarate dehydrogenase*), 6.2.1.4 (*Succinyl coenzyme A synthetase*), 1.3.5.1 (*Succinate dehydrogenase*), 4.2.1.2 (*Fumarate hydratase*), 1.1.1.37 (*Malate dehydrogenase*). There are two full reactions (metabolite, protein and transcript level) This figure was adapted from KEGG reference pathway.

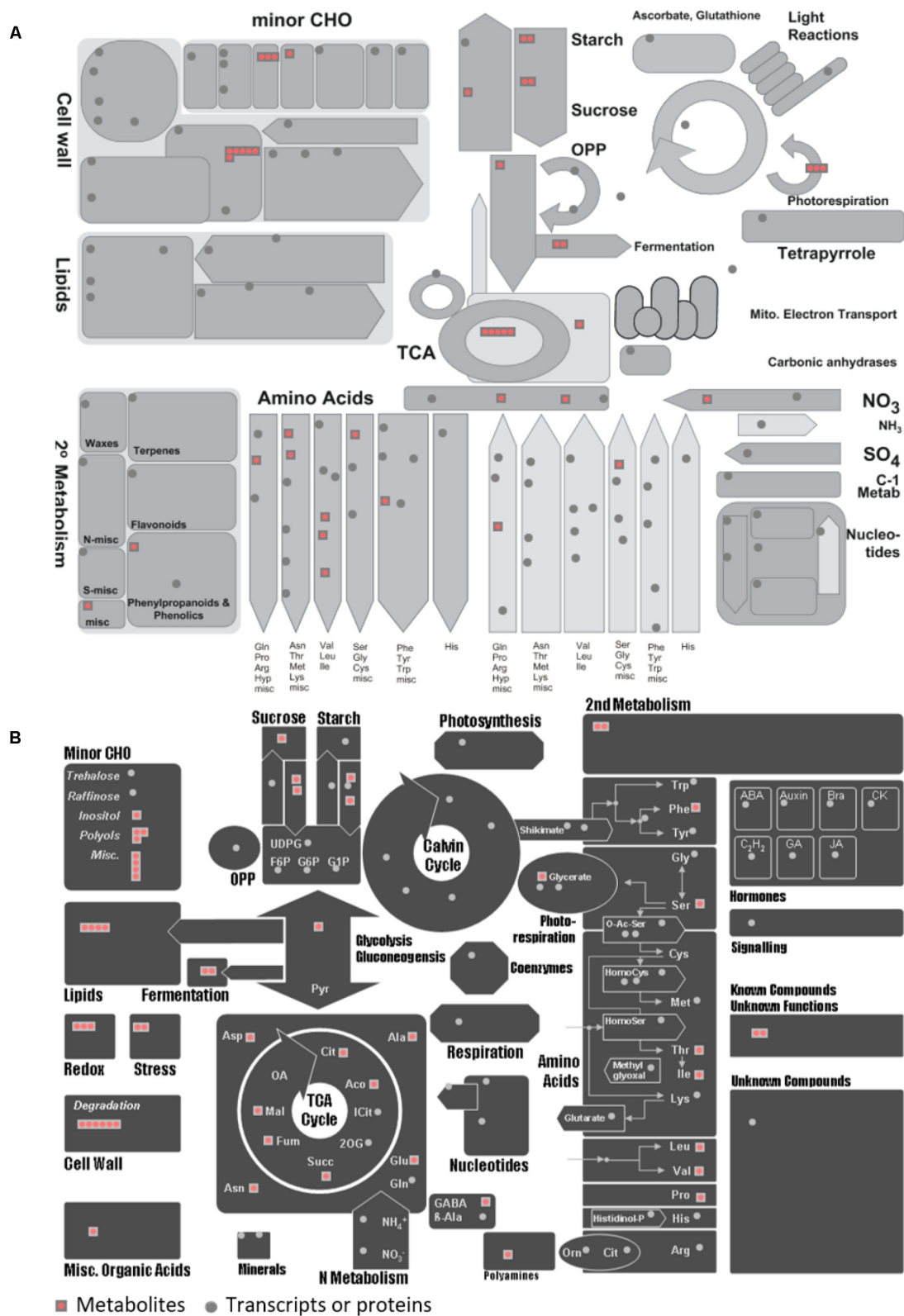
functional plant categories according to MapMan classification (**Figure 1**).

These metabolites are starting metabolites or final products from primary metabolism pathways, like glyoxylate and dicarboxylate metabolism (*ath00630*), starch and sucrose metabolism (*ath00500*), citrate cycle (TCA cycle) (*ath00020*) of carbohydrate metabolism; alanine, aspartate, and glutamate metabolism (*ath00250*) of amino acid metabolism and biosynthesis of unsaturated fatty acids (*ath01040*) of fatty acids metabolism. Many were intermediate metabolites, with 5 (citrate, *cis*-Aconitate, succinate, fumarate, and malate), out of the total 8 corresponding to the Citrate cycle (**Figure 4** and **Table 2**). The pathways most represented were carbohydrate and amino acid metabolisms. However, the number of secondary metabolites (catechin, epigallocatechin, and anthraquinone) was smaller than the number of secondary metabolites reported for *Quercus* spp. acorns (Vinha et al., 2016a). Due to the small number of secondary metabolites detected, the metabolic pathways related to the biosynthesis of secondary metabolites, like carotenoid

biosynthesis (*ath00906*), anthocyanin biosynthesis (*ath00942*), and monoterpenoid biosynthesis (*ath00902*) are not highly represented (**Supplementary Table S3**). In *Arabidopsis*, the total number of secondary metabolites is still unknown due to metabolite identification being one of the bottlenecks in untargeted metabolomic studies (Wu et al., 2017). Still, in AraCyc 15.0, the total number of compounds described are 2971 and the number of metabolic pathways 610 (PMN; Plant Metabolic Network<sup>18</sup>).

The identification of 62 metabolites is in the order of what has been reported for non-model plant systems by using a similar approach (Warren et al., 2012; Cadahia et al., 2014; Asai et al., 2016; Pascual et al., 2017), but far from the figure obtained when using model systems such as *A. thaliana*, or complementary techniques such as LC-MS. The employment of complementary LC-MS strategies would increase the number of metabolites identified, as shown, for example, with *A. thaliana*, although it would greatly reduce

<sup>18</sup><https://www.plantcyc.org/>



**FIGURE 5 |** MapMan overview of general metabolism for the metabolites and proteins/transcripts of *Quercus ilex*. **(A)** Visualization of 58 metabolites in the context of general metabolism using the MapMan software. **(B)** Visualization of 58 metabolites in different MapMan pathways. Each red square represents a metabolite and each gray circle represents a protein or transcript. More details can be found in Usadel et al. (2009).

the number of metabolites identified with no doubts. Kim et al. (2015) detected 4483 distinct metabolite peaks from leaves using 11 mass spectrometric platforms, but only identifying 1348 metabolites. These results revealed that the available databases and repositories are incomplete and pointed to the need for new algorithms for elucidating structures from MS<sup>n</sup> analyses.

## Data Integration

To seek insights into the metabolic pathways as they occur in Holm oak, transcriptomics, proteomics, and metabolomics data have been integrated. (Table 1 and Supplementary Tables S8, S10). We obtained a deeper view of the metabolic pathways by implementing proteomics or transcriptomics data as the potential of these techniques is much higher than that of metabolomics. However, although technological advances and bioinformatic tools and resources for making those analyses and data interpretation have been extended to plant biology research, this has mostly been for model plants. The unique and specialized biology of such diversified species requires the adaptation of strategies conceived primarily for model organisms and the development of designed and specific methods. For their integration, we employed EC numbers (proteins and transcripts) and KEGG identifiers (metabolites). With the latter and with KEGG pathway maps we obtained the three-different level of information of 61 metabolic pathways (Supplementary Table S3). The metabolic pathways most represented are shown in Table 2. In order to obtain a metabolic overview. The “BINS” generated from the proteome/transcriptome were employed as a “mapping file,” then introducing identified metabolites. The representation obtained of the general map (Figure 5A) for the dataset as shown from ImageAnnotator module of MapMan, showed common metabolism points between metabolites and proteins/transcripts (Figure 5B). From the total number of pathways reported in the plants, for example, in KEGG (127 pathways in *Arabidopsis*), we procured data from 124 of them at the metabolomic, proteomic, and transcriptomics level (Supplementary Figure S3). Table 2 summarizes the most representative pathways visualized, including carbohydrate metabolism [glycolysis/gluconeogenesis (*ath00010*), glyoxylate and dicarboxylate metabolism (*ath00630*), citrate cycle (TCA cycle) (*ath00020*), starch and sucrose metabolism (*ath00500*)], amino acid metabolism [alanine, aspartate, and glutamate metabolism (*ath00250*) and phenylalanine metabolism (*ath00360*), lipid metabolism (biosynthesis of unsaturated fatty acids (*ath01040*)), and energy metabolism [carbon fixation in photosynthetic organisms (*ath00710*)]. The one most represented was the TCA, with 5 metabolites out of a total of 10, and protein and transcript corresponding to, respectively, 6 and 8 enzymes (Figure 2). On the other hand, there were clear gaps in the hypothetical plant metabolic chart, mainly corresponding to the secondary metabolism and hormones [anthocyanin biosynthesis (*ath00942*), brassinosteroid biosynthesis (*ath00905*)] and lipid metabolism [steroid biosynthesis (*ath00100*)]. For example, the brassinosteroid biosynthesis pathway, which produces plant steroidal hormones that play important roles in many stages of plant growth, has only reported 1 protein and 1

transcript (Supplementary Table S4). Also, the Figure 3 shows the low representation of the different metabolic pathways, also with a multi-omics data integration. From metabolomics, proteomics, and transcriptomic data we were able to identify 64, 109, and 118, pathways, respectively. The total number reported at the PMN and deduced from genome sequencing were 610 (*A. thaliana*), 519 (*E. grandis*), and 538 (*P. trichocarpa*). From these figures we can conclude that the current wet methodologies only allow the visualization of a low percentage of enzyme gene products in a single experiment.

The work and dataset generated, even considering future methodological improvements, will be the basis of ulterior studies on the particularities of the metabolism as it occurs in different organs and developmental processes, as well the changes in response to environmental cues, thus complementing our previous studies in which morphology, phenology, classical physiological and biochemical analysis, and the holistic proteomics have been employed (Echevarría-Zomeño et al., 2009, 2012; Valero-Galván et al., 2011, 2012a; Sghaier-Hammami et al., 2013; Romero-Rodríguez et al., 2014, 2015; Romero-Rodríguez, 2015; Simova-Stoilova et al., 2015; Guerrero-Sanchez et al., 2017). These previously published studies provided quite fragmented and speculative biological information. Hence to go one step ahead, data validation and integration at the different molecular levels would be necessary in order to obtain an unbiased molecular interpretation of the plant biology.

## CONCLUSION

We have proven that – omics integration, in the Systems Biology direction, is feasible not only with model organisms, but also with orphan and recalcitrant species such as the Holm oak, the most emblematic and representative tree species of the Mediterranean forest. The methodological bases, including wet protocols and *in silico* analysis, have been established, allowing the implementation of transcriptome, proteome, and metabolome databases, comprising 27089 transcripts (unigenes), 2380 protein species, and 62 metabolites (Supplementary Table S11).

Integrated analysis allowed the visualization and reconstruction of the metabolism in Holm oak. Up to 123 metabolic pathways, out of the 127-total reported in KEGG, can be visualized at the transcriptome, proteome, and metabolome level. Thus, as an example, for the Krebs cycle, six metabolites out of the eight have been detected. This route comprises eight enzymes detected at the transcriptome or proteome level. These figures are like those reported for the model plant *A. thaliana*. There is still room for improvement, and there are pathways underrepresented in the created database, including the brassinosteroid biosynthesis pathway. The *Q. ilex* genome sequencing, the use of alternative and complementary strategies such as LC-MS will improve the number of pathways visualized.

The current metabolic reconstruction achieved for this species can be considered to be sufficient to progress in the biological knowledge of this species.

## DATA AVAILABILITY

RAW and MSF files corresponding to proteomics are available at the ProteomExchange repository; Datasets: PXD008001. The Project ID of the GC-MS *Q. ilex* metabolomic analysis is PR000618 in the Metabolomics Workbench repository.

## AUTHOR CONTRIBUTIONS

CL-H performed the GC/MS experiments, analyzed the data, and wrote the manuscript. VG-S analyzed the data and wrote the manuscript. IG-G performed the proteomics experiments. RS-L performed the proteomics experiments. MC-S performed the proteomics experiments and wrote the manuscript. AM-A performed the transcriptomics experiments. LV supervised and wrote the manuscript. JJ-N conceived and designed the experiments, supervised and wrote the manuscript.

## FUNDING

This work was supported by the University of Córdoba and financial support from the Spanish Ministry of Economy and Competitiveness (Project BIO2015-64737-R2). The staff of the Central Service for Research Support (SCAI) at the University of Córdoba is acknowledged for its technical support in metabolomics and proteomics data analysis.

## ACKNOWLEDGMENTS

We are grateful to Centro Informático Científico de Andalucía (Spain) for helping us to make the transcriptome annotations. Proyecto CICYT.

## REFERENCES

- Abril, N., Gion, J. M., Kerner, R., Müller-Starck, G., Cerrillo, R. M. N., Plomion, C., et al. (2011). Proteomics research on forest trees, the most recalcitrant and orphan plant species. *Phytochemistry* 72, 1219–1242. doi: 10.1016/j.phytochem.2011.01.005
- Akcan, T., Gökçe, R., Asensio, M., Estévez, M., and Morcuende, D. (2017). Acorn (*Quercus* spp.) as a novel source of oleic acid and tocopherols for livestock and humans: discrimination of selected species from Mediterranean forest. *J. Food Sci. Technol.* 54, 3050–3057. doi: 10.1007/s13197-017-2740-3
- Al-Rousan, W. M., Ajo, R. Y., Al-Ismail, K. M., Attlee, A., Shaker, R. R., and Osaili, T. M. (2013). Characterization of acorn fruit oils extracted from selected Mediterranean *Quercus* species. *Grasas Aceites* 64, 554–560. doi: 10.3989/gya.023313
- Asai, T., Matsukawa, T., and Kajiyama, S. (2016). Metabolomic analysis of primary metabolites in citrus leaf during defense responses. *J. Biosci. Bioeng.* 123, 376–381. doi: 10.1016/j.jbiosc.2016.09.013
- 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
- Cadahía, E., De Simón, B. F., Aranda, I., Sanz, M., Sánchez-Gómez, D., and Pinto, E. (2014). Non-targeted metabolomic profile of *Fagus sylvatica* L. leaves using liquid chromatography with mass spectrometry and gas chromatography with mass spectrometry. *Phytochem. Anal.* 26, 171–182. doi: 10.1002/pca.2549

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.00935/full#supplementary-material>

**FIGURE S1** | Density histogram for proteins in the different biological processes of *Q. ilex* annotated transcriptome.

**FIGURE S2** | Density histogram for proteins in the different cellular components of *Q. ilex* annotated transcriptome.

**FIGURE S3** | Density histogram for proteins in the different molecular functions of *Q. ilex* annotated transcriptome.

**FIGURE S4** | Enzymes (transcript level and protein level) assigned to the glycolysis/gluconeogenesis.

**TABLE S1** | Metabolite features.

**TABLE S2** | GC-MS metabolomic data. Mean values of normalized peak areas, as well as SD (standard deviation) and CV% (percentage of coefficient of variation) were determined for replicates of the metabolite extract.

**TABLE S3** | KEGG pathways with metabolites, proteins, and transcripts.

**TABLE S4** | Comparison of KEGG pathways of different species.

**TABLE S5** | Bins of transcripts.

**TABLE S6** | Bins of proteins.

**TABLE S7** | Comparison of *in silico* proteomes.

**TABLE S8** | Shotgun LC-MS/MS proteomic data. Mean values of normalized peak areas, as well as SD (standard deviation) and CV% (percentage of coefficient of variation) were determined for replicates of protein extract (Jorge et al., 2005, 2006). The area values correspond to replicate 1 (0.6 µg of protein), replicate 2 (0.8 µg of protein), and replicate 3 (1 µg of protein).

**TABLE S9** | Enzymes (transcripts).

**TABLE S10** | Enzymes (proteins).

**TABLE S11** | Omics overview.

- Cantos, E., Espín, J. C., López-Bote, C., De la Hoz, L. D., Ordóñez, J. A., and Tomás-Barberán, F. A. (2003). Phenolic compounds and fatty acids from acorns (*Quercus* spp.), the main dietary constituent of free-ranged Iberian pigs. *J. Agric. Food Chem.* 51, 6248–6255. doi: 10.1021/jf030216v
- Casimiro-Soriguer, C. S., Muñoz-Mérida, A., and Pérez-Pulido, A. J. (2017). Sma3s: a universal tool for easy functional annotation of proteomes and transcriptomes. *Proteomics* 17, 1–12. doi: 10.1002/pmic.201700071
- Castillejo, M. Á., Bani, M., and Rubiales, D. (2015). Understanding pea resistance mechanisms in response to *Fusarium oxysporum* through proteomic analysis. *Phytochemistry* 115, 44–58. doi: 10.1016/j.phytochem.2015.01.009
- Chen, H., and Boutros, P. C. (2011). VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* 12:35. doi: 10.1186/1471-2105-12-35
- Chevreur, B., and Suhai, S. (1999). Genome sequence assembly using trace signals and additional sequence information. *Ger. Conf. Bioinform.* 99, 45–56.
- Chira, K., and Teissedre, P. L. (2014). Chemical and sensory evaluation of wine matured in oak barrel: effect of oak species involved and toasting process. *Eur. Food Res. Technol.* 240, 533–547. doi: 10.1007/s00217-014-2352-3
- Correia, B., Valledor, L., Hancock, R. D., Renaut, J., Pascual, J., Soares, A. M. V. M., et al. (2016). Integrated proteomics and metabolomics to unlock global and clonal responses of *Eucalyptus globulus* recovery from water deficit. *Metabolomics* 12:141. doi: 10.1007/s11306-016-1088-4
- De Wit, P., Pespeni, M. H., Ladner, J. T., Barshis, D. J., Seneca, F., Jaris, H., et al. (2012). The simple fool's guide to population genomics via RNA-Seq: an



- introduction to high-throughput sequencing data analysis. *Mol. Ecol. Resour.* 12, 1058–1067. doi: 10.1111/1755-0998.12003
- Echevarría-Zomeño, S., Abril, N., Ruiz-Laguna, J., Jorrín-Novo, J., and Maldonado-Alconada, A. M. (2012). Simple, rapid and reliable methods to obtain high quality RNA and genomic DNA from *Quercus ilex* L. leaves suitable for molecular biology studies. *Acta Physiol. Plant.* 34, 793–805. doi: 10.1007/s11738-011-0880-z
- Echevarría-Zomeño, S., Ariza, D., Jorge, I., Lenz, C., Del Campo, A., Jorrín, J. V., et al. (2009). Changes in the protein profile of *Quercus ilex* leaves in response to drought stress and recovery. *J. Plant Physiol.* 166, 233–245. doi: 10.1016/j.jplph.2008.05.008
- Escandón, M., Valledor, L., Pascual, J., Pinto, G., Casal, M. J., and Meijón, M. (2017). System-wide analysis of short-term response to high temperature in *Pinus radiata*. *J. Exp. Bot.* 68, 3629–3641. doi: 10.1093/jxb/erx198
- Furuhashi, T., Fragner, L., Furuhashi, K., Valledor, L., Sun, X., and Weckwerth, W. (2012). Metabolite changes with induction of *Cuscuta* haustorium and translocation from host plants. *J. Plant Interact.* 7, 84–93. doi: 10.1080/17429145.2011.603059
- Guerrero-Sanchez, V. M., Maldonado-Alconada, A. M., Amil-Ruiz, F., and Jorrín-Novo, J. V. (2017). Holm oak (*Quercus ilex*) transcriptome. De novo sequencing and assembly analysis. *Front. Mol. Biosci.* 4:70. doi: 10.3389/fmolb.2017.00070
- Hadidi, L., Babou, L., Zaidi, F., Valentão, P., Andrade, P. B., and Grosso, C. (2017). *Quercus ilex* L.: how season, plant organ and extraction procedure can influence chemistry and bioactivities. *Chem. Biodivers.* 14:e1600187. doi: 10.1002/cbdv.201600187
- Hoagland, D. R., and Arnon, D. I. (1950). *The Water-Culture Method for Growing Plants without Soil*. California Agricultural Experiment Station, Circular-347. Available at: <http://hdl.handle.net/2027/uc2.ark:/13960/t51g1sb8j>
- Janda, M., Planchais, S., Djafi, N., Martinec, J., Burketova, L., Valentova, O., et al. (2013). Phosphoglycerolipids are master players in plant hormone signal transduction. *Plant Cell Rep.* 32, 839–851. doi: 10.1007/s00299-013-1399-0
- Jorge, I., Navarro, R. M., Lenz, C., Ariza, D., and Jorrín, J. (2006). Variation in the Holm oak leaf proteome at different plant developmental stages, between provenances and in response to drought stress. *Proteomics* 6, S207–S214. doi: 10.1002/pmic.200500364
- Jorge, I., Navarro, R. M., Lenz, C., Ariza, D., Porras, C., and Jorrín, J. (2005). The holm oak leaf proteome: analytical and biological variability in the protein expression level assessed by 2-DE and protein identification tandem mass spectrometry de novo sequencing and sequence similarity searching. *Proteomics* 5, 222–234. doi: 10.1002/pmic.200400893
- Jorrín-Novo, J. V., Komatsu, S., Weckwerth, W., and Wienkoop, S. (eds). (2014). “Analysis of rice proteins using SDS-PAGE shotgun proteomics,” in *Plant Proteomics: Methods and Protocols* (New York, NY: Humana Press Inc.), 1072.
- Kalli, A., Smith, G. T., Sweredoski, M. J., and Hess, S. (2013). Evaluation and optimization of mass spectrometric settings during data-dependent acquisition mode: focus on LTQ- orbitrap mass analyzers. *J. Proteome Res.* 12, 3071–3086. doi: 10.1021/pr3011588
- Kamboj, J. S., Blake, P. S., Quinlan, J. D., and Baker, D. A. (1999). Identification and quantitation by GC-MS of zeatin and zeatin riboside in xylem sap from rootstock and scion of grafted apple trees. *Plant Growth Regul.* 28, 199–205. doi: 10.1023/A:1006292309765
- Kim, T., Dreher, K., Nilo-Poyanco, R., Lee, I., Fiehn, O., Lange, B. M., et al. (2015). Patterns of metabolite changes identified from large-scale gene perturbations in *Arabidopsis* using a genome-scale metabolic network. *Plant Physiol.* 167, 1685–1698. doi: 10.1104/pp.114.252361
- Kind, T., Wohlgemuth, G., Lee, D. Y., Palazoglu, M., Shahbaz, S., et al. (2010). FiehnLib – mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Anal. Chem.* 81, 10038–10048. doi: 10.1021/ac9019522.fiehnlib
- Li, W., Jaroszewski, L., and Godzik, A. (2001). Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics* 17, 282–283. doi: 10.1093/bioinformatics/bth457
- Li, W., Jaroszewski, L., and Godzik, A. (2002). Tolerating some redundancy significantly speeds up clustering of large protein databases. *Bioinformatics* 18, 77–82. doi: 10.1093/bioinformatics/18.1.77
- Lohse, M., Nagel, A., Herter, T., May, P., Schroda, M., Zrenner, R., et al. (2014). Mercator: a fast and simple web server for genome scale functional annotation of plant sequence data. *Plant Cell Environ.* 37, 1250–1258. doi: 10.1111/pce.12231
- López-Hidalgo, C. (2017). *Análisis Metabolómico de Especies Forestales, la Encina (Quercus ilex) (Final Master Project)*. Córdoba: Universidad de Córdoba.
- Meijón, M., Feito, I., Oravec, M., Delatorre, C., Weckwerth, W., Majada, J., et al. (2016). Exploring natural variation of *Pinus pinaster* Aiton using metabolomics: is it possible to identify the region of origin of a pine from its metabolites? *Mol. Ecol.* 25, 959–976. doi: 10.1111/mec.13525
- Muñoz-Mérida, A., Viguera, E., Claros, M. G., Trelles, O., and Pérez-Pulido, A. J. (2014). Sma3s: a three-step modular annotator for large sequence datasets. *DNA Res.* 21, 341–353. doi: 10.1093/dnares/dsu001
- Nielsen, J., and Jewett, M. (2007). *Metabolomics: A Powerful Tool in Systems Biology*. Berlin: Springer. doi: 10.1007/978-3-540-74719-2
- Pascual, J., Cañal, M. J., Escandón, M., Meijón, M., Weckwerth, W., and Valledor, L. (2017). Integrated physiological, proteomic, and metabolomic analysis of ultra violet (UV) stress responses and adaptation mechanisms in *Pinus radiata*. *Mol. Cell. Proteomics* 16, 485–501. doi: 10.1074/mcp.M116.059436
- Pluskal, T., Castillo, S., Villar-Briones, A., and Orešič, M. (2010). MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* 11:395. doi: 10.1186/1471-2105-11-395
- Rabhi, F., Narváez-Rivas, M., Tlili, N., Boukhchina, S., and León-Camacho, M. (2016). Sterol, aliphatic alcohol and tocopherol contents of *Quercus ilex* and *Quercus suber* from different regions. *Ind. Crops Prod.* 83, 781–786. doi: 10.1016/j.indcrop.2015.11.020
- Rakić, S., Povrenović, D., Tešević, V., Simić, M., and Maletić, R. (2006). Oak acorn, polyphenols and antioxidant activity in functional food. *J. Food Eng.* 74, 416–423. doi: 10.1016/j.jfoodeng.2005.03.057
- Rennenberg, H., Herschbach, C., Haberer, K., and Kopriva, S. (2007). Sulfur metabolism in plants: are trees different? *Plant Biol.* 9, 620–627. doi: 10.1055/s-2007-965248
- Rice, P., Longden, I., and Bleasby, A. (2000). EMBOS: the European molecular biology open software suite. *Trends Genet.* 16, 276–277. doi: 10.1016/S0168-9525(00)00204-2
- Rigo, D., and De Caudullo, G. (2016). “*Quercus ilex* in Europe?: distribution, habitat, usage and threats,” in *European Atlas of Forest Tree Species*, eds J. San-Miguel-Ayán, D. de Rigo, G. Caudullo, T. Houston Durrant, A. Mauri (Brussels: European Commission), 152–153.
- Romero-Rodríguez, M. C. (2015). *Integrated “Omics” Approaches to Study Non - Orthodox Seed Germination: The Case of Holm Oak (Quercus ilex subsp. ballota [Desf.] Samp)*. Doctoral thesis, University of Córdoba, Córdoba, 234.
- Romero-Rodríguez, M. C., Abril, N., Sánchez-Lucas, R., and Jorrín-Novo, J. V. (2015). Multiplex staining of 2-DE gels for an initial phosphoproteome analysis of germinating seeds and early grown seedlings from a non-orthodox specie: *Quercus ilex* L. subsp. *ballota* [Desf.] Samp. *Front. Plant Sci.* 6:620. doi: 10.3389/fpls.2015.00620
- Romero-Rodríguez, M. C., Pascual, J., Valledor, L., and Jorrín-Novo, J. V. (2014). Improving the quality of protein identification in non-model species. Characterization of *Quercus ilex* seed and *Pinus radiata* needle proteomes by using SEQUEST and custom databases. *J. Proteomics* 105, 85–91. doi: 10.1016/j.jprot.2014.01.027
- Sardans, J., Gargallo-Garriga, A., Pérez-Trujillo, M., Parella, T. J., Seco, R., Filella, I., et al. (2014). Metabolic responses of *Quercus ilex* seedlings to wounding analysed with nuclear magnetic resonance profiling. *Plant Biol.* 16, 395–403. doi: 10.1111/plb.12032
- Schrimpe-Rutledge, A. C., Codreanu, S. G., Sherrod, S. D., and McLean, J. A. (2016). Untargeted metabolomics strategies-challenges and emerging directions. *J. Am. Soc. Mass Spectrom.* 27, 1897–1905. doi: 10.1007/s13361-016-1469-y
- Sghaier-Hammami, B., Valero-Galván, J., Romero-Rodríguez, M. C., Navarro-Cerrillo, R. M., Abdelly, C., and Jorrín-Novo, J. V. (2013). Physiological and proteomics analyses of Holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.) responses to *Phytophthora cinnamomi*. *Plant Physiol. Biochem.* 71, 191–202. doi: 10.1016/j.plaphy.2013.06.030
- Simova-Stoilova, L. P., Romero-Rodríguez, M. C., Sánchez-Lucas, R., Navarro-Cerrillo, R. M., Medina-Aunon, J. A., and Jorrín-Novo, J. V. (2015). 2-DE proteomics analysis of drought treated seedlings of *Quercus ilex* supports a root

- active strategy for metabolic adaptation in response to water shortage. *Front. Plant Sci.* 6:627. doi: 10.3389/fpls.2015.00627
- Szuba, A., and Lorenc-Plucińska, G. (2017). Field proteomics of *Populus alba* grown in a heavily modified environment – An example of a tannery waste landfill. *Sci. Total Environ.* 610, 1557–1571. doi: 10.1016/j.scitotenv.2017.06.102
- Tarkka, M. T., Herrmann, S., Wubet, T., Feldhahn, L., Recht, S., Kurth, F., et al. (2013). OakContigDF159.1, a reference library for studying differential gene expression in *Quercus robur* during controlled biotic interactions: use for quantitative transcriptomic profiling of oak roots in ectomycorrhizal symbiosis. *New Phytol.* 199, 529–540. doi: 10.1111/nph.12317
- The Angiosperm Phylogeny Group III (2009). An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* 161, 105–121. doi: 10.1111/j.1095-8339.2009.00996.x
- Tian, H., Lam, S. M., and Shui, G. (2016). Metabolomics, a powerful tool for agricultural research. *Int. J. Mol. Sci.* 17:E1871. doi: 10.3390/ijms17111871
- Usadel, B., Poree, F., Nagel, A., Lohse, M., Czédik-Eysenberg, A., and Stitt, M. (2009). A guide to using MapMan to visualize and compare Omics data in plants: a case study in the crop species, Maize. *Plant Cell Environ.* 32, 1211–1229. doi: 10.1111/j.1365-3040.2009.01978.x
- Valero-Galván, J., Jorrín-Novo, J. V., Gómez Cabrera, A., Ariza, D., García-Olmo, J., and Navarro-Cerrillo, R. M. (2012a). Population variability based on the morphometry and chemical composition of the acorn in Holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.). *Eur. J. For. Res.* 131, 893–904. doi: 10.1007/s10342-011-0563-8
- Valero-Galván, J., Valledor, L., Cerrillo, R. M. N., Pelegrín, E. G., and Jorrín-Novo, J. V. (2011). Studies of variability in Holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.) through acorn protein profile analysis. *J. Proteomics* 74, 1244–1255. doi: 10.1016/j.jprot.2011.05.003
- Valero-Galván, J., Valledor, L., González Fernandez, R., Navarro Cerrillo, R. M., and Jorrín-Novo, J. V. (2012b). Proteomic analysis of Holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.) pollen. *J. Proteomics* 75, 2736–2744. doi: 10.1016/j.jprot.2012.03.035
- Valledor, L., Escandón, M., Meijón, M., Nukarinen, E., Cañal, M. J., and Weckwerth, W. (2014). A universal protocol for the combined isolation of metabolites, DNA, long RNAs, small RNAs, and proteins from plants and microorganisms. *Plant J.* 79, 173–180. doi: 10.1111/tbj.12546
- Viant, M. R., Kurland, I. J., Jones, M. R., and Dunn, W. B. (2017). How close are we to complete annotation of metabolomes? *Curr. Opin. Chem. Biol.* 36, 64–69. doi: 10.1016/j.cbpa.2017.01.001
- Vinha, A. F., Barreira, J. C. M., Costa, A. S. G., and Oliveira, M. B. P. P. (2016a). A new age for *Quercus* spp. fruits: review on nutritional and phytochemical composition and related biological activities of acorns. *Compr. Rev. Food Sci. Food Saf.* 15, 947–981. doi: 10.1111/1541-4337.12220
- Vinha, A. F., Costa, A. S. G., Barreira, J. C. M., Pacheco, R., and Oliveira, M. B. P. P. (2016b). Chemical and antioxidant profiles of acorn tissues from *Quercus* spp.: potential as new industrial raw materials. *Ind. Crops Prod.* 94, 143–151. doi: 10.1016/j.indcrop.2016.08.027
- Warren, C. R., Aranda, I., and Cano, F. J. (2012). Metabolomics demonstrates divergent responses of two *Eucalyptus* species to water stress. *Metabolomics* 8, 186–200. doi: 10.1007/s11306-011-0299-y
- Wu, S., Tohge, T., Cuadros-Inostroza, Á., Tong, H., Tenenboim, H., Kooke, R., et al. (2017). Mapping the Arabidopsis metabolic landscape by untargeted metabolomics at different environmental conditions. *Mol. Plant* 11, 118–134. doi: 10.1016/j.molp.2017.08.012
- Zhang, R., Calixto, C. P. G., Marquez, Y., Venhuizen, P., Tzioutziou, N. A., Guo, W., et al. (2017). A high quality Arabidopsis transcriptome for accurate transcript-level analysis of alternative splicing. *Nucleic Acids Res.* 45, 5061–5073. doi: 10.1093/nar/gkx267

**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 López-Hidalgo, Guerrero-Sánchez, Gómez-Gálvez, Sánchez-Lucas, Castillejo-Sánchez, Maldonado-Alconada, Valledor and Jorrín-Novo. 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.



# Metabolite Profiles of Sugarcane Culm Reveal the Relationship Among Metabolism and Axillary Bud Outgrowth in Genetically Related Sugarcane Commercial Cultivars

## OPEN ACCESS

### Edited by:

Glória Catarina Pinto,  
University of Aveiro, Portugal

### Reviewed by:

Antonio Figueira,  
Universidade de São Paulo, Brazil  
Frikkie C. Botha,  
Sugar Research Australia, Australia

### \*Correspondence:

Camila Caldana  
caldana@mpimp-golm.mpg.de

### †Present address:

Danilo A. Ferreira,  
Bayer CropScience, São Paulo, Brazil  
Adriana Cheavegatti-Gianotto,  
Centro de Tecnologia Canavieira,  
Piracicaba, Brazil  
Camila Caldana,  
Max Planck Institute of Molecular  
Plant Physiology, Potsdam, Germany

‡These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Plant Systems and Synthetic Biology,  
a section of the journal  
Frontiers in Plant Science

Received: 03 November 2017

Accepted: 01 June 2018

Published: 25 June 2018

### Citation:

Ferreira DA, Martins MCM,  
Cheavegatti-Gianotto A, Carneiro MS,  
Amadeu RR, Aricetti JA, Wolf LD,  
Hoffmann HP, de Abreu LGF and  
Caldana C (2018) Metabolite Profiles  
of Sugarcane Culm Reveal  
the Relationship Among Metabolism  
and Axillary Bud Outgrowth  
in Genetically Related Sugarcane  
Commercial Cultivars.  
Front. Plant Sci. 9:857.  
doi: 10.3389/fpls.2018.00857

**Danilo A. Ferreira<sup>1,2†</sup>, Marina C. M. Martins<sup>1†</sup>, Adriana Cheavegatti-Gianotto<sup>1†</sup>,  
Monalisa S. Carneiro<sup>3</sup>, Rodrigo R. Amadeu<sup>4</sup>, Juliana A. Aricetti<sup>1</sup>, Lucia D. Wolf<sup>1</sup>,  
Hermann P. Hoffmann<sup>3</sup>, Luis G. F. de Abreu<sup>1</sup> and Camila Caldana<sup>1,5\*</sup>**

<sup>1</sup> Brazilian Bioethanol Science and Technology Laboratory, Centro Nacional de Pesquisa em Energia e Materiais, Campinas, Brazil, <sup>2</sup> Genetics and Molecular Biology Graduate Program, University of Campinas, Campinas, Brazil, <sup>3</sup> Department of Biotechnology and Plant and Animal Production, Center for Agricultural Sciences, Federal University of São Carlos, São Carlos, Brazil, <sup>4</sup> Department of Genetics, Luiz de Queiroz College of Agriculture, University of São Paulo, Piracicaba, Brazil, <sup>5</sup> Max-Planck Partner Group, Brazilian Bioethanol Science and Technology Laboratory, Centro Nacional de Pesquisa em Energia e Materiais, Campinas, Brazil

Metabolic composition is known to exert influence on several important agronomic traits, and metabolomics, which represents the chemical composition in a cell, has long been recognized as a powerful tool for bridging phenotype–genotype interactions. In this work, sixteen truly representative sugarcane Brazilian varieties were selected to explore the metabolic networks in buds and culms, the tissues involved in the vegetative propagation of this species. Due to the fact that bud sprouting is a key trait determining crop establishment in the field, the sprouting potential among the genotypes was evaluated. The use of partial least square discriminant analysis indicated only mild differences on bud outgrowth potential under controlled environmental conditions. However, primary metabolite profiling provided information on the variability of metabolic features even under a narrow genetic background, typical for modern sugarcane cultivars. Metabolite–metabolite correlations within and between tissues revealed more complex patterns for culms in relation to buds, and enabled the recognition of key metabolites (e.g., sucrose, putrescine, glutamate, serine, and myo-inositol) affecting sprouting ability. Finally, those results were associated with the genetic background of each cultivar, showing that metabolites can be potentially used as indicators for the genetic background.

**Keywords: sugarcane, breeding, metabolome, bud outgrowth, metabolic network**

## INTRODUCTION

Plants have an extraordinarily complex metabolism, and a comprehensive understanding on how it operates pose a challenge due to the coordination among various biochemical processes in specialized tissues and subcellular compartments (Lunn, 2007; Sweetlove and Fernie, 2013). Their sessile nature adds an extra layer of difficulty, as there is a constant need to adjust to changes in the

surrounding environment (Jaillais and Chory, 2010; Kooke and Keurentjes, 2012). This significant level of organization allows the production of a plethora of chemical compounds, which differ in their properties (e.g., size, polarity, stability, and quantity) and biological functions, representing a readout of the physiological status of a cell. Traditionally, plant metabolomics studies have focused on elucidating the function and regulation of particular biosynthetic routes involving a number of metabolites (Stitt et al., 2010; Fernie and Tohge, 2017). However, advances in large-scale automated analytical platforms have increasingly enabled high-throughput detection of metabolites, allowing the elucidation of metabolic networks in terms of structure and connectivity and/or bridging the genotype-to-phenotype gap to elucidate certain biological processes. Although knowledge about the role of specific enzymes was extended by targeted reverse genetics approaches (Alex et al., 2004; Tohge et al., 2005; Zheng et al., 2005; Seki et al., 2011; Goulet et al., 2015), duplication of enzymes and their different subcellular localization hampers metabolic engineering modifications relying on a single transgenic (Huang et al., 2010; Qin et al., 2011; Ren et al., 2014; Fernie and Tohge, 2017). Means to surpass this problem include the use of natural/genetic variance to enhance our understanding about the genetic architecture of metabolic traits and monitor gene-phenotype combinations in a wide range of plant species (e.g., *Arabidopsis*, tomato, and rice) or important agronomic traits such as fruit composition (Bernillon et al., 2013; Monti et al., 2016), grain yield (Obata et al., 2015; Dan et al., 2016), and tolerance to abiotic stresses (Glaubitz et al., 2014; Todaka et al., 2017; Sprenger et al., 2018).

As metabolism is strongly influenced by interactions between the environment and genetic regulation, there is a limitation to extrapolate the complete picture of plant metabolomes by evaluating a single condition (i.e., developmental stage, genetic background and environment) (Soltis and Kliebenstein, 2015). Furthermore, apart from having their biosynthesis and accumulation in a tissue-specific manner, metabolites can be produced and transported across tissues and/or organs to mediate certain biological processes. One example of this kind of regulation is the fate of axillary buds, which is governed by a complex interplay among environmental factors, genetic background and endogenous metabolites (Huang et al., 2012). Metabolite signals arising from other parts of the plant such as shoots or stems are sensed prior to trigger systemic responses that will promote bud outgrowth (Dun et al., 2012; Barbier F.F. et al., 2015; Brewer et al., 2015). Several hormones have been long recognized as the main signaling molecules in this process (Umehara et al., 2008; Domagalska and Leyser, 2011; Durbak et al., 2012). However, the availability of sugars, especially sucrose, was recently found to be crucial for bud outgrowth release prior to alterations in hormone levels (Mason et al., 2014; Barbier F. et al., 2015). Manipulation of sucrose supply via decapitation or defoliation was able to promote or suppress bud outgrowth, respectively (Mason et al., 2014; Barbier F. et al., 2015; Fichtner et al., 2017). Interestingly, dormant buds present a transcriptional response related to carbon starvation that seems to be conserved among different species (Tarancón et al., 2017). Primary metabolites, such as sugars and amino acids, are integral

parts of sophisticated signaling networks linking the energetic status and external cues to regulate growth accordingly (Nunes-Nesi et al., 2010; Smeekens et al., 2010; Xiong et al., 2013; Chellamuthu et al., 2014; Yadav et al., 2014). Collectively, this new information placed primary metabolites as essential molecules with more immediate roles in bud development and outgrowth.

The regulation of axillary bud outgrowth is crucial for crops in which either vegetative propagation or tillering are important traits, as it is the case of the perennial C4 grass sugarcane (*Saccharum × officinarum*). In sugarcane, axillary buds are also naturally in a dormant state (Jain et al., 2009), however, when segments of the culms containing portions of internode and node with embryo roots and at least one viable bud are isolated from the plant body and placed into soil, bud outgrowth is released and a new plant is generated. Sugarcane is capable of accumulating impressive amounts of sucrose in its stems, in a very complex and dynamic process characterized by a continuous cycle of synthesis and degradation (Whittaker and Botha, 1997; Zhu et al., 1997; Botha and Black, 2000; Rose and Botha, 2000), which involves various enzymes and their isoforms. There is a gradient of sucrose accumulation along the stem, with younger internodes containing less sucrose than older internodes (Zhu et al., 1997; Pereira et al., 2017). Interestingly, the stored carbon in the form of sucrose is used for bud outgrowth and formation of a new sugarcane plant (O'Neill et al., 2012). Due to the fact that sucrose was shown to be crucial for bud outgrowth in other species (Mason et al., 2014; Barbier F. et al., 2015; Kebrom and Mullet, 2015), it remains to be elucidated whether the remarkably high levels of sucrose or other components of the primary metabolism are important to promote bud outgrowth release in sugarcane. The complex genetic architecture of sugarcane (e.g., high polyploidy, high heterozygosity, large amount of repetitive sequences, aneuploidy, and large genome size) (Zhang et al., 2012) has hampered the use of genetic information to dissect biological mechanisms in this species (de Setta et al., 2014; Song et al., 2016; Riaño-Pachón and Mattiello, 2017; Hoang et al., 2017). All these characteristics make the application of metabolomics a great alternative for investigating complex agronomic traits such as sprouting potential.

In the present study, we assessed the metabolic profile of two tissues involved in the sprouting potential, namely culm and bud, from 16 highly planted sugarcane varieties from a Brazilian sugarcane breeding program (varietal census 2016/2017<sup>1</sup>). The cultivars studied herein rank among the most cultivated genotypes in the world as they cover about 65% of the sugarcane planted area in Brazil, the major sugarcane producer worldwide. These cultivars are therefore a worthy sample of sugarcane commercial genotypes with greater field performance. Our results demonstrate that the culm metabolism plays an important role as primary energy source to provide carbon skeletons and building blocks for protein synthesis for triggering bud outgrowth. Overall, our results suggest that both factors, genetic background and bud sprouting rates, jointly influenced the metabolite profile of sugarcane, opening perspectives for the use of metabolomics to assist sugarcane breeding programs.

<sup>1</sup><https://www.ridesa.com.br/censo-varietal>



## MATERIALS AND METHODS

### Plant Material

A collection of 16 relevant genotypes was chosen from the leading Brazilian sugarcane-breeding program The Inter-University Network for the Development of Sugarcane Industry (RIDESA) (**Supplementary Table S1**). Out of 16, the varieties RB867515, RB966928 and RB92579 cover 42% of total sugarcane fields in the country (2016/2017 varieties census: see footnote text 1). Sugarcane breeding programs have been indirect selected genotypes with high sprouting potential by choosing experimental plots with higher density and stalk yield. Consequently, the current commercial cultivars have low variability to this trait and tend to present medium to high sprouting potential under field conditions (Cargnin et al., 2008). The sugarcane breeding programs rely generally on a limited number of elite plant material as parental lines. Therefore, these 16 selected genotypes were also used as a proof of concept to evaluate whether even under a narrow genetic basis metabolic profiles could be used to discriminate their metabolic status. Information on the parents of the selected genotypes was recorded from RIDESA database, the pedigree tree (**Supplementary Figure S1**) was drawn using the R packages *synbreed* (Wimmer et al., 2012) and *diagram* (Soetaert, 2017). The degree of kinship among the cultivars is represented as the coefficient of relationship between the individuals (Wright, 1922) computed using *AGHmatrix* package (Amadeu et al., 2016; **Supplementary Figure S2**).

### Experimental Conditions

Field and greenhouse experiments were conducted at the Federal University of São Carlos (UFSCar)/RIDESA in Araras, São Paulo, Brazil located at 22°21'25" S, 47°23'03" W, about 611 m above sea level; in a typic eutroferic red latosol soil. Mature sugarcane plants (approximately 11 months old) were decapitated 24 h before the harvest in the field, to facilitate the loss of apical dominance. After that, three stems of each genotype from independent plants were randomly harvested around 2 h after dawn. For each stem, internodes were counted and divided into three parts. Due to variations in sprouting performance throughout the stem according to bud position, only internodes belonging to the middle portion of the stem were further cut close to the bud, and had their diameter and weight measured to guarantee uniformity. Considering the existence of a sucrose gradient as well as different developmental stages along the sugarcane stem, the selection of the middle third portion of this organ would allow a better comparison among the genotypes. It is worth mentioning that usually entire sugarcane stems are planted in commercial field environments.

This material, also known as setts, was used for both metabolite profiling analysis and sprouting performance evaluation. In case of metabolite profiling, buds and the region of the culm in which they were inserted in were precisely isolated with the help of a scalpel and cork borer, respectively. A total of 3 biological replicates (representing three independent stalks), each containing a pool of three individual buds or culms, was

collected. After the harvesting process, which took approximately 5 min per genotype, tissues were immediately frozen into liquid nitrogen and stored at  $-80^{\circ}\text{C}$  for metabolic profiling analysis.

Setts were planted with buds oriented toward the light into 200 ml pots containing commercial substrate Plantmax® for sprouting evaluation. Since sugarcane initial development is sensitive to soil water content and changes in temperature (Oliveira et al., 2001; Singels and Smit, 2002; Smit, 2011), the experiment was performed in the greenhouse during May 2016 with automated irrigation system (six times along the diel cycle) and the temperature was set to 35 and 29°C before and after sprouting, respectively. Each genotype was planted in three completely randomized trays containing 24 individuals each.

### Sprouting Rate Evaluation

Sprouting performance was assessed in the greenhouse by monitoring bud outgrowth during the first 14 days after planting. Even considering the lack of synchronization in bud outgrowth release and their potential to be viable over longer terms, dormant buds would hardly become seedlings under field conditions after this period. Sprouting was considered successful when the seedling stem crossed the soil surface (a layer of 2 cm of soil over the bud) and was able to issue the first leaf. In order to classify the genotypes according to the sprouting rate, a descriptive quartile analysis was performed, in which varieties belonging to the top or down 25% of data distribution were considered with high or low sprouting potential, respectively. To further improve the classification of the genotypes belonging to the middle 50% quartile of data distribution, a second quartile analysis was performed to distinguish them as intermediate-low or -high sprouting potential. For assessing the variances, a parametric test (*F*-test) was applied to all genotypes at 5% of significance levels.

### Metabolite Profiling Analysis

Prior to metabolite profiling analyses, sugarcane tissues were ground to a fine powder in liquid nitrogen and aliquots of 20 or 50 mg for culms and buds, respectively, were used for metabolite extraction, following the methodology described by Giavalisco et al. (2011). A fraction of 100  $\mu\text{l}$  from the organic phase was dried and derivatized as described in Roessner et al., 2001. Afterwards, 1  $\mu\text{l}$  of the derivatized samples was analyzed on a Combi-PAL autosampler (Agilent Technologies GmbH, Waldbronn, Germany) coupled to an Agilent 7890 gas chromatograph and a Leco Pegasus 2 time-of-flight mass spectrometer (LECO, St. Joseph, MI, United States) in both split (1:40 and 1:65 for buds and culm, respectively) and splitless modes (Weckwerth et al., 2004). Chromatograms were exported from Leco ChromaTOF software (version 3.25) to R software. Peak detection, retention time alignment, and library matching using the Golm Database<sup>2</sup> were performed using TargetSearch R package (Cuadros-Inostroza et al., 2009). Metabolites were quantified by the peak intensity of a selective mass. Metabolites intensities were normalized by dividing the fresh weight of each

<sup>2</sup><http://gmd.mpimp-golm.mpg.de/>

biological replicate, followed by the sum of total ion count and log<sub>2</sub> transformation.

## Statistical Analyses

Statistical analyses and graphical representations were performed using R version 3.2.3<sup>3</sup>. Multivariate analyses, including PCA and PLS-DA, were carried out using mixOmics (Rohart et al., 2017) and pcaMethods (Stacklies et al., 2007) R packages. Correlation analysis, heatmap and network visualization were done using corrrplot (Wei and Simko, 2017), d3heatmap (Cheng et al., 2018), and qgraph (Epskamp et al., 2012) packages, respectively.

## RESULTS

### Commercial Sugarcane Varieties Displayed Mild to Low Variability in Bud Outgrowth Under Controlled Environmental Conditions

During the initial period of seedling establishment (14 days), an overall high sprouting homogeneity was observed among genotypes. The first quartile analysis allowed the classification of 31, 56, and 13% of the genotypes into low, intermediate, and high sprouting potential, respectively. To refine this classification, a second quartile analysis was performed only using the intermediate genotypes. However, due to the fairly homogenous sprouting ability of the genotypes observed in the present study, the subdivision of this group resulted only in 19 and 38% of genotypes with intermediate-low and intermediate-high sprouting potential, respectively. Out of the 16 selected genotypes, RB975375 and RB935744 were classified as high sprouting rate, whereas RB937570, RB975201, RB835486, RB966928, and RB72454 were considered with low sprouting potential (Supplementary Figure S3). Despite these major groups, the sprouting rate was in the range of 89–100% and no statistical differences among the genotypes were observed at level of 5% of significance. Due to the nature of vegetative propagation of this crop, breeding programs have indirectly favored genotypes with high sprouting rates. However, there is still a severe lack of synchronization of the bud outgrowth in the Brazilian field conditions, especially in areas susceptible to drought, leading to massive reduction in sprouting and consequently jeopardizing yield (Cargnin et al., 2008). Furthermore, the sprouting rate estimation was based on the establishment of a new plant, and although this is the measured trait in the field, which indirectly reflects the bud outgrowth performance, it does not enable the assessment of the internal factors involved in the control of bud release.

### Metabolite Profiling Revealed Differential Metabolic Responses of Genotypes in Distinct Tissues

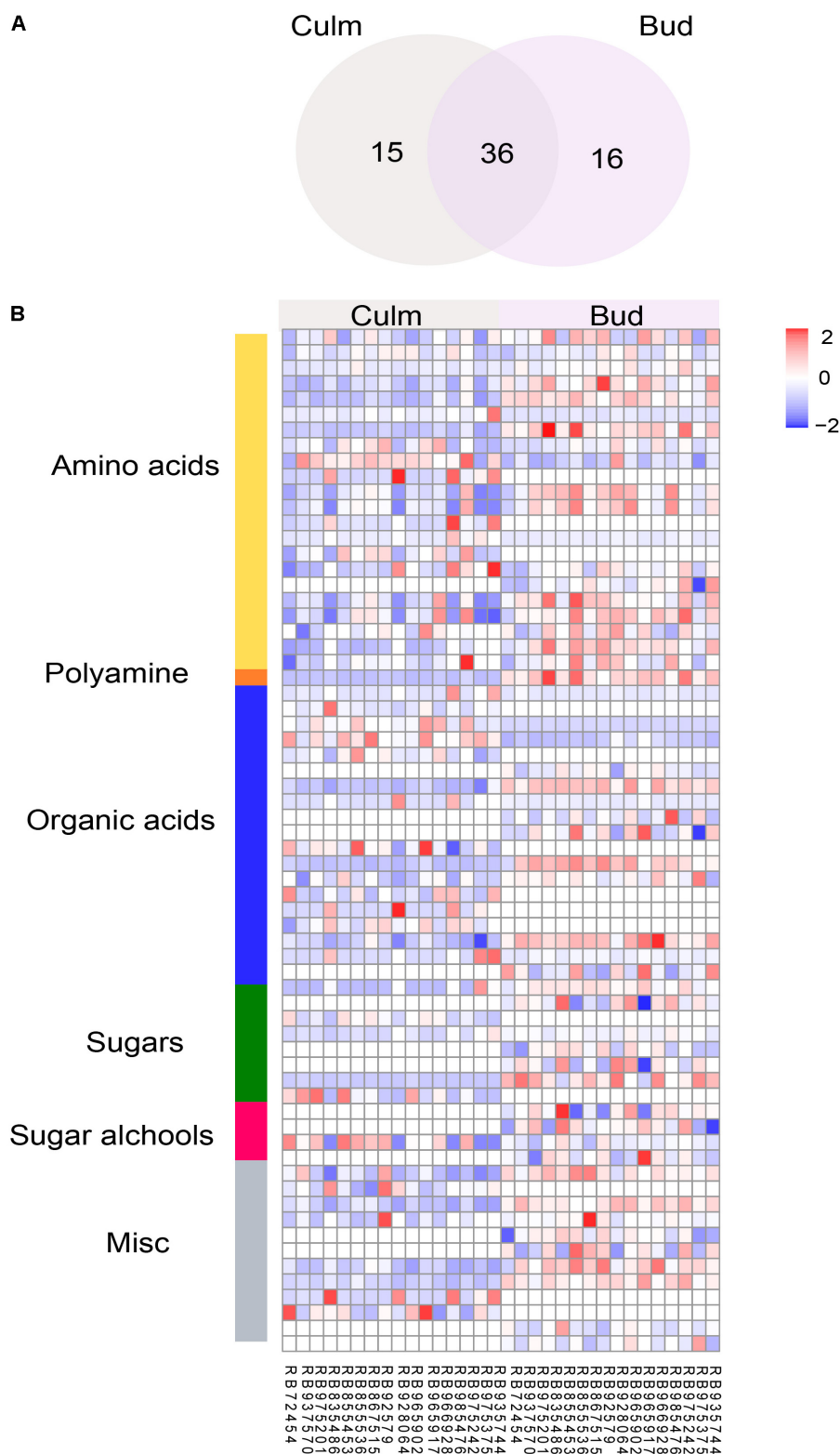
As it is challenging to morphologically and molecularly monitor the factors triggering bud outgrowth release in this crop, we

next investigated whether the metabolite profiling could be a great tool to understand the control of bud outgrowth using GC-MS. This platform allows the assessment of molecules involved mainly in central metabolism, which was already reported to be closely linked to plant growth (Meyer et al., 2007; Lisec et al., 2008). Due to the fact that no significant differences were found in water content of the studied genotypes (data not shown), we used fresh weight for normalizing the metabolite levels. A major portion of the metabolites (76.7 %) was found in all samples. Due to the saturation of sucrose levels, the same samples were also injected in a split mode (diluted 1:40 or 1:65 for buds and culms, respectively) to accurately quantify this sugar. We detected a total of 66 metabolites with known structures (e.g., amino acids, sugar, sugar alcohols, organic acids, and polyamines), of which 16 and 15 were specifically identified in bud and culm, respectively (Figure 1A). Figure 1B shows a heatmap including the metabolite abundance of genotypes in each tissue and Supplementary Table S2 summarizes the effect of individual metabolites. By applying ANOVA, we found that most metabolites were significantly affected by the genotype in both tissues (Supplementary Figure S4), indicating enough variability in metabolic features among genotypes even under a narrow genetic basis (Supplementary Figures S1, S2).

### Metabolite–Metabolite Correlations Provide New Insights on the Regulation of Metabolic Networks Intra- and Inter-Tissues

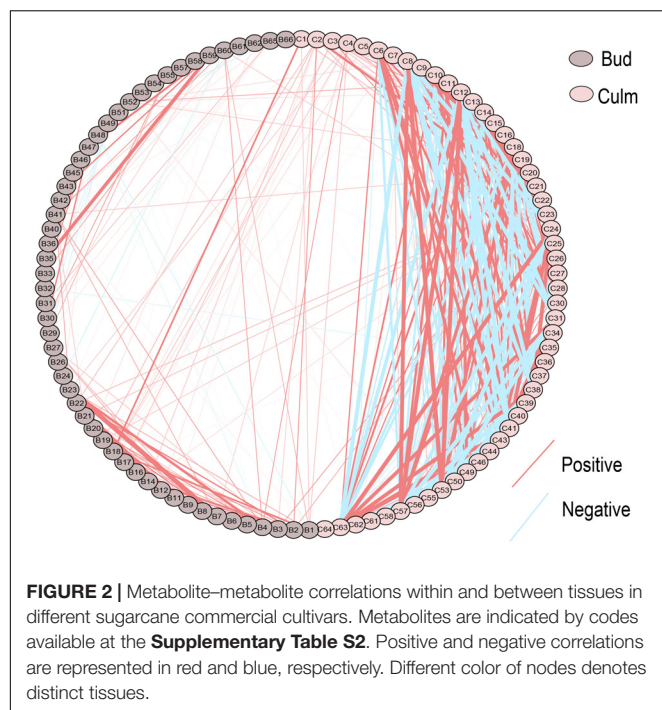
To decipher the relationships among metabolites, we performed correlation-based network analysis using significant pairwise correlations ( $r \geq 0.5$ ,  $p \leq 0.05$ ) (Figure 2), which are summarized in Supplementary Table S3. As expected, metabolites belonging to the same biochemical pathway tended to present a high degree of connectivity as it was the case for valine, isoleucine and leucine ( $r > 0.8$ ) in both tissues. In total, there were 148, 414, and 47 significant correlations among metabolites detected in buds, culms and between these two tissues, respectively, suggesting that the metabolite–metabolite correlations were diverse in the different tissues. An exception was a subnetwork containing positive correlations among the branched amino acids leucine, isoleucine and valine, and threonine, conserved in both tissues. Interestingly, those amino acids were only linked to each other and methionine in buds, whereas in culm this network became more complex. Apart from the branched amino acids and threonine, further highly positive connections were built among glutamine, serine and the sugars sucrose, fructose and myo-inositol in culms. This expanded subnetwork was negatively linked to another subnetwork including GABA, putrescine, benzoate, galacturonate, nicotinate, and a metabolite with similarity to itaconate, which in turn were all positively correlated to each other. The strongest negative correlation ( $r = -0.9679$ ,  $p = 0.05$ ) was between sucrose and putrescine that is one of the links between these two subnetworks in culms. Remarkably, the role of those two metabolites in controlling bud dormancy have been recently shown (Cui et al., 2010; Mason et al., 2014; Barbier F. et al., 2015). Interestingly, the

<sup>3</sup><https://www.r-project.org/>



**FIGURE 1 |** Metabolic composition of buds and culms in different sugarcane commercial cultivars. **(A)** Number of metabolites identified in culm, bud, and common to both tissues; **(B)** Differential abundance of metabolites in culms and buds according to their compound class. Each column represents the genotype mean, and each row represents a metabolite. Variations in the relative abundance of metabolites are displayed in blue (low) to red (high).





compound similar to itaconate, which displayed the strongest positive correlation in culm with galacturonate ( $r = 0.977$ ,  $p = 0.05$ ) presented distinct correlations with amino acids in culms and buds. One example is the connection with glutamate that seems to be the opposite in both tissues. Glutamate and serine are the only metabolites that significant connect the bud and culm network. With respect to the bud, its network is much less interconnected when compared to culm. In addition to the amino acids subnetwork, another highly interconnected subnetwork containing galactose, quinate and sorbose ( $r > 0.83$ ,  $p = 0.05$ ) was identified. Altogether, these results revealed that the metabolic network of culm is more coordinately regulated than that of bud. A logical explanation is that culms constitute more active tissues with pivotal role as strong sinks, not only related to sucrose, but also including amino acids, which will act as primary energy source to provide carbon skeletons and building blocks for protein synthesis during bud outgrowth.

## Culm Metabolic Composition Is Important for Bud Outgrowth Performance

We also investigated whether the metabolic composition of both tissues among the selected commercial cultivars would have an impact on sprouting even under low trait variability. Based on the role of polyamines and sugars controlling axillary meristem dormancy and tillering/branching (Zheng et al., 2005; Ge et al., 2006; Purohit et al., 2007; Cui et al., 2010; Mason et al., 2014; Barbier F. et al., 2015), we hypothesized that the metabolic composition of the culm could be one of the key factors determining bud outgrowth. Although our analysis was restricted

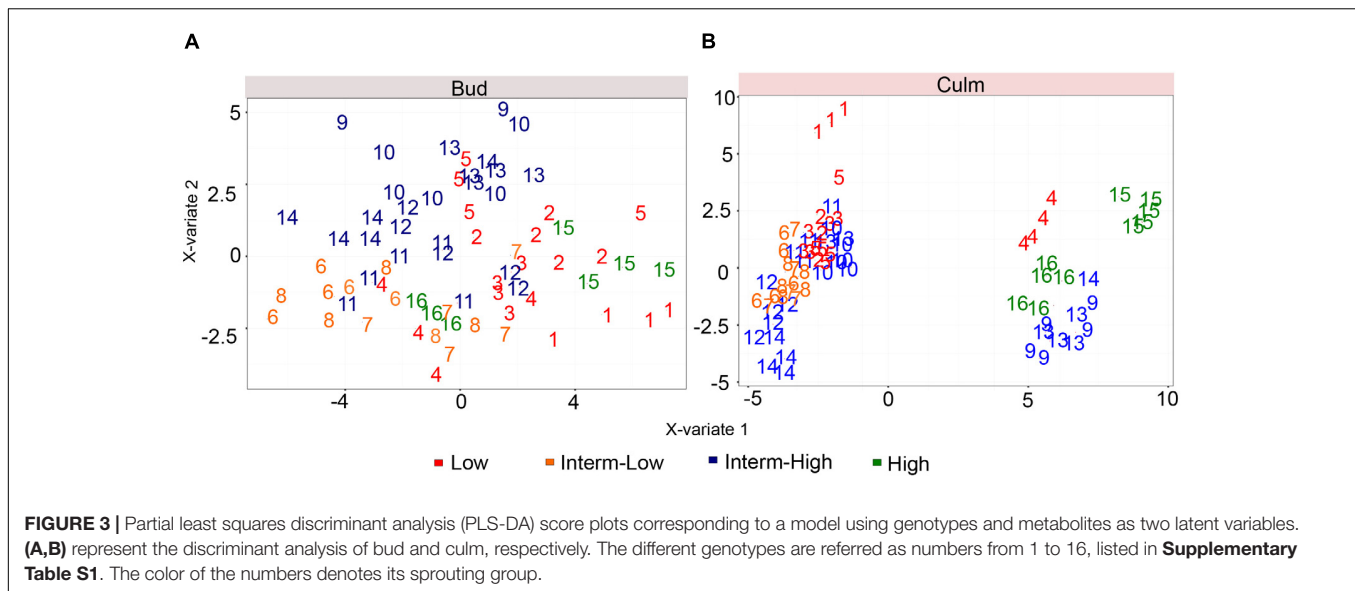
to few compounds of the primary metabolism, it covers key metabolites known to play fundamental roles in promoting sink to source transitions and plant growth. A partial least squares discriminant analysis (PLS-DA) was applied in an attempt to understand the role of culms in bud outgrowth (**Figure 3**). Our results showed that the bud metabolomic data solely did not result in a good separation among genotypes according to the sprouting potential (**Figure 3A**). One plausible reason is that different genotypes might be at distinct stages of dormancy, which could exert influence on their metabolism hampering the discrimination based on the metabolome. Furthermore, as the metabolic activities of mature culms are committed to sucrose storage and reduced in comparison to fast growing stages of the plant life cycle, we cannot exclude the possibility that changes in metabolic contents among cultivars are partially masked at this specific phase. In contrast to bud, the PLS-DA revealed a significant difference in the culm metabolism for low and high sprouting rate at least for the most contrasting genotypes (**Figure 3B**).

To identify the metabolites responsible for the separation among genotypes, a cluster image analysis was performed by building a similarity matrix with the PLS-DA results (**Figure 4**). A total of five well-defined groups was obtained and although there was no clear association in three of the clusters with respect to sprouting rate, two groups presented a very clear trend in relation to bud outgrowth performance. Interestingly, the levels of several metabolites including lysine, histidine, phenylalanine, GABA, methionine, xylose, benzoate, tetradecanoate, putrescine, nicotinate, galacturonate, the compound similar to itaconate, putrescine, nicotinate, benzoate, and galacturonate seemed to correlate positively with the high sprouting rate genotypes. In contrast, the levels of glycerol, quinate, fructose, sucrose, myo-inositol, leucine, glutamine, isocitrate, glutamate, ornithine, serine, threonine, and isoleucine appeared to have a negative impact on bud outgrowth. Taken together, our results showed that the culm, but not the bud metabolome, enabled the discrimination of the genotypes based on their sprouting performance only among the most contrasting cultivars. Furthermore, the culm metabolism turned up important to determine bud outgrowth efficiency as shown by the correlation of contrasting genotypes with antagonistic metabolites such as sugars and polyamines, this latter suggested as a signaling mediator in bud dormancy, and also by the presence of glutamate and serine, both involved in the connection of culm and bud networks.

## Sugarcane Metabolome Reflects Sprouting Rate and the Genetic Relatedness of Commercial Genotypes

As the culm metabolome permitted to rank at some extent the genotypes according to the sprouting rate, we next investigated how central metabolism was influenced by the genetic background of the selected cultivars in the conditions of this experiment. For that, we compared the hierarchical clustering analysis (HCA) considering only the culm metabolome (**Figure 5**) to numerator relationship matrix





(**Supplementary Figure S2**) obtained from pedigree information shown in **Supplementary Figure S1**.

The HCA analysis revealed the presence of two defined clusters (**Figure 5**). The first cluster represents a very similar group of eleven genotypes with low to moderate-high sprouting rates. Interestingly, pairs of individuals with a relationship coefficient of 0.5 (parent-offspring or full-sib) in the numerator relationship matrix (**Supplementary Figure S2**) tended to be kept in this HCA cluster. This cluster was mainly composed of individuals genetically related to the genotype RB72454, which was used as parental of several crossings of this panel (**Supplementary Figure S1**). The mismatch individual, RB975375, displayed 100% of sprouting rate. The other exception was the genotype RB835486, which clustered together with RB72454 despite their complete absence of relatedness as per pedigree information (**Supplementary Figure S1**). Interesting to note that RB835486 displayed the third lowest sprouting rate (93.1%), which main explain its placement at this first cluster.

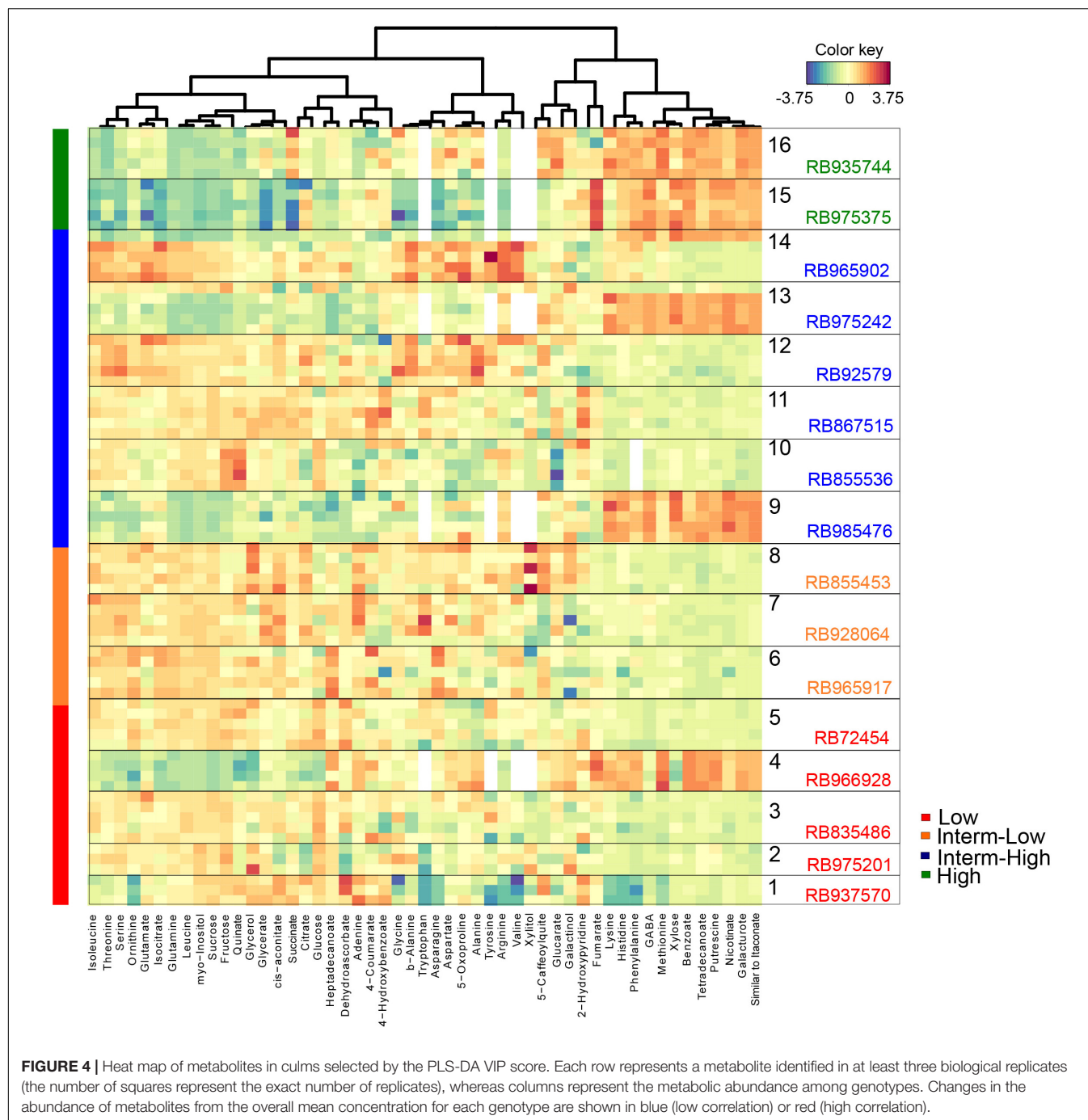
The second cluster represents cultivars with high and moderate-high sprouting rates (RB935744 and RB975375 – high; RB985476 and RB975242 – moderate-high) with the exception of cultivar RB966928 that presents low sprouting rate. All the cultivars located in this cluster tended to display parentage coefficient below 0.25, among themselves and with other cultivars of this panel (**Supplementary Figure S2**), according to the information from our pedigree. This fact is reflected in the height of the HCA dendrogram, which confirms their relative weaker genetic relationship.

It is worth to mention that our pedigree estimate is incomplete due the lack of information about the fathers of cultivars obtained from multiparental crosses. Besides, the estimation of the kinship matrix (**Supplementary Figure S2**) could also be improved using informative molecular markers, which were out of the scope of this work. Despite this, the numerator relationship matrix (**Supplementary Figure S2**) revealed that the parentage coefficient was already high among cultivars (average

of 0.114), suggesting a close relationship among cultivars (i.e., such value is almost the coefficient between cousins, 0.125). Apart from RB92579 and RB975242, all the clones share some degree of relatedness with at least one individual in the panel (**Supplementary Figures S1, S2**). Overall, our results suggest that both factors, genetic background and bud sprouting rates, jointly influenced the metabolomic profile of sugarcane revealed by HCA when evaluated under a single environmental condition.

## DISCUSSION

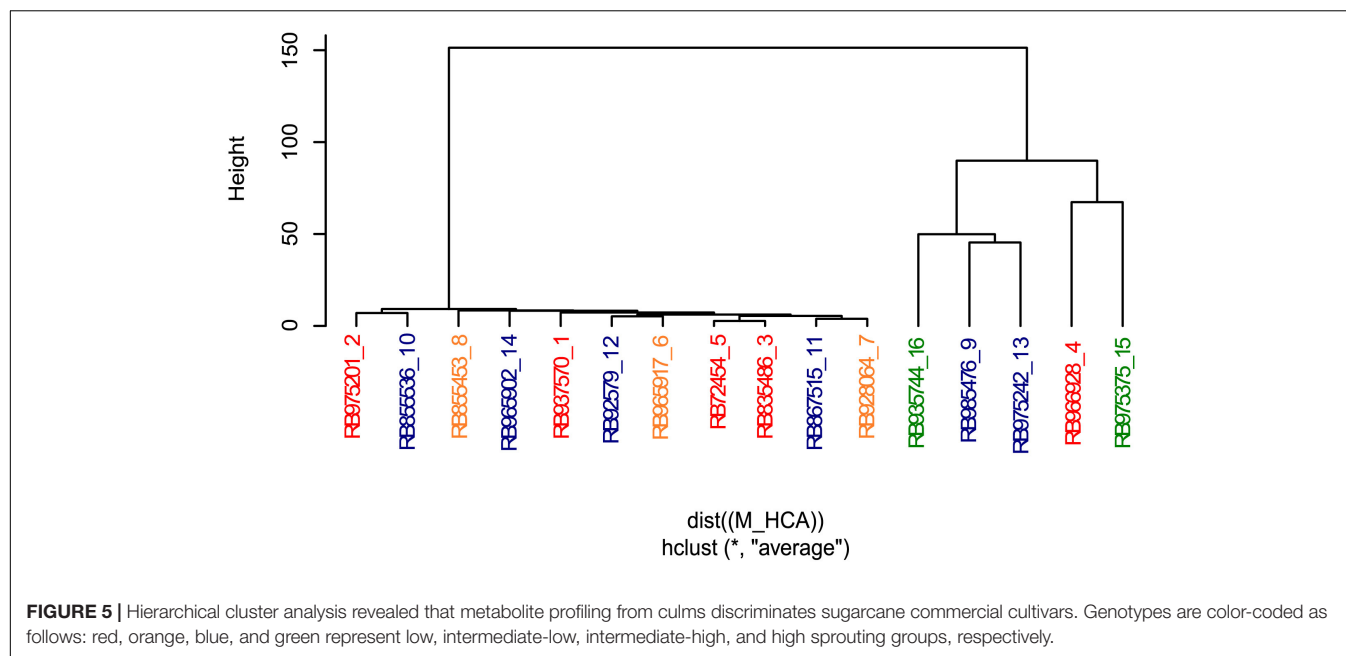
Metabolomics has been widely used as a powerful tool to elucidate mechanisms involved in metabolic regulation as well as bridging the gap between genotype and phenotype (Saito and Matsuda, 2010; Kusano et al., 2011; Kumar et al., 2017). Over the last decade, metabolomics has been also used in association with natural variation to unravel the genetic architecture of several agronomic traits controlled by metabolism (Meyer et al., 2007; Toubiana et al., 2012; Witt et al., 2012). Although these studies have provided many insights into biological properties, the complexity of the metabolome and its dependency on the environment, genetics and development, precludes the generation of a full complete picture (Soltis and Kliebenstein, 2015). In this study, we minimized this complexity by fixing the environmental conditions to assess whether the primary metabolism of two organs involved in the sprouting potential is regulated to trigger bud outgrowth in 16-highly planted sugarcane cultivars, representing a good sampling of commercial cultivars with overall good field performance, including good sprouting rate. Under our experimental conditions (controlled temperature and water availability), the selected sugarcane genotypes presented low variability in their sprouting rate (about 10%). Optimal sprouting conditions rarely occur in important production areas, when buds are often exposed to several abiotic constraints. One example is the Brazilian central region, in



which the impact of limiting environmental conditions resulted in sprouting rates ranging from 29,17 to 76,92% among 8 commercial cultivars. This work evaluated 4 cultivars (RB867515, RB855453, RB835486, RB855536) also studied herein, but only RB855453 and RB855536 presented sprouting ratios over 70% under those environmental conditions (Cargnin et al., 2008). Furthermore, it is important to mention that a mix of internodes from top, middle and bottom portions of the stem is planted in commercial fields. In our experimental setup, we selected only internodes belonging to the middle part of the stem to minimize

the variation in sprouting rate dependent on developmental stage and bud position among the varieties (Manhães et al., 2015; Baracat-Neto et al., 2017).

It is widely known that there is a gradient concentration of sucrose along the sugarcane stem, with mature internodes having higher sucrose levels that decrease toward the top immature internodes. Strikingly, the bud sprouting performance along the stem follows the opposite gradient of sucrose (Whittaker and Botha, 1997; Zhu et al., 1997; Vorster and Botha, 1999; Uys et al., 2007) and the culm metabolism is apparently essential



to determine the dormant status of the axillary meristem and bud outgrowth (Boussiengui-Boussiengui et al., 2016). Although few studies aimed to investigate the mechanisms responsible for successful germination/sprouting in this species (Verma et al., 2013; Singh et al., 2016; Boussiengui-Boussiengui et al., 2016), very little is known about the biochemical and molecular aspects related to this process, especially concerning sink and source interactions of the bud and culm. Plant growth and development is modulated by the balance between source and sink strengths (Paul and Foyer, 2001; Dingkuhn et al., 2007; Smith and Stitt, 2007; Patrick and Colyvas, 2014), namely production of photoassimilates in leaves and their use in non-photosynthetic organs. In sugarcane, sucrose can be quickly metabolized in sink tissues to maintain its levels within a proper range, enabling fast responses to alterations in sucrose supply and demand. However, depending on the carbon demand, the culm starts to act as an additional source tissue, mobilizing sucrose to sustain developmental transitions. In the case of sprouting, sucrose will be used to promote axillary bud outgrowth and seedling establishment (O'Neill et al., 2012). The amino acids leucine and isoleucine seem also to play a later role during this process (O'Neill et al., 2012) and isotopic analysis demonstrated that nitrogen reserves from the culms are important for seedling establishment in the first 50–60 days of development (Carneiro et al., 1995). In this sense, remobilization of carbon and nitrogen mediated by sucrose and amino acids from the culm is crucial for the establishment of a new shoot from the axillary meristem. However, it still remains to be elucidated which key metabolites participate in breaking the axillary meristem dormancy. It is known that bud outgrowth is inhibited by the action of hormones in a phenomenon termed apical dominance, which can be suppressed by excision, developmental transitions or diseases (Botha et al., 2013; Rameau et al., 2015; Barbier et al., 2017). In this work we focused on investigating how primary

compounds of central metabolism, rather than hormones, behave and interact in buds and culms during sugarcane bud outgrowth. Unraveling the interaction between hormone and central metabolite signaling will be crucial to dissect the temporal cascade controlling bud outgrowth release.

As plant growth regulation is closely modulated by the primary metabolism, our study used GC-MS-based metabolomics to unravel these relationships in culm and bud among different genotypes. Due to the fact that the modern commercial sugarcane varieties have narrow genetic basis, we first addressed if metabolic features would display any degree of variability among the selected sugarcane cultivars. Statistical tests (ANOVA) on metabolome data did not only confirm metabolic variability, but also unravel differences between the studied organs.

In order to further unravel small molecules involved in this process, metabolite–metabolite correlation analysis was performed within and between tissues. Several studies pinpointed the high connectivity of amino acids in Arabidopsis, tomato and maize (Schauer et al., 2006; Toubiana et al., 2015; Wen et al., 2015), suggesting that their network is controlled by a high degree of metabolic regulation (Galili and Höfgen, 2002). Accordingly, our results showed that overall amino acids were highly correlated. Interestingly, glutamate and serine were among the few metabolites presenting correlations between culm and bud. Glutamate, a hub in amino acid metabolism, is substrate of glutamine synthetase (GS) to generate glutamine or is formed by the conversion of glutamine and 2-oxoglutarate in the presence of either reduced ferredoxin (Fd) or NADH by glutamate synthase (GOGAT) during inorganic ammonium assimilation (Lea and Miflin, 1974; Yamaya and Oaks, 2004). In rice, transgenic plants lacking the cytosolic glutamine synthetase 1;2 (GS1;2) exhibited a severe suppression of bud outgrowth (Ohashi et al., 2015), suggesting a role of glutamate as signal molecule for sensing nitrogen status and controlling this process. Furthermore,

glutamate is also a precursor of serine biosynthesis in a non-photorespiratory route called phosphorylated pathway, which was the other metabolite linking culm to bud metabolism and has been shown to control cell proliferation (Cascales-Minana et al., 2013; Ros et al., 2014). In this context, metabolomics is a powerful tool for identifying candidate metabolic pathways involved in diverse biological processes.

With respect to the tissue-specific metabolic networks, the culm presented a more coordinately regulated metabolism than the bud. Due to its high concentration in parenchyma cells of stem internodes, sucrose was one of the main hubs in the culm network, as expected. This disaccharide was responsible for the most negative correlation in the network with the putrescine, an important precursor for polyamine biosynthesis. Polyamines are aliphatic nitrogen compounds that have been proposed to be involved in many processes during plant growth and development in response to environmental cues (Kusano et al., 2007; Gill and Tuteja, 2010) and are crucial for plant survival as blockage of their biosynthesis leads to lethal phenotypes (Urano et al., 2005; Ge et al., 2006). Interestingly, deletion in one of the genes encoding for the enzyme S-adenosylmethionine decarboxylases, involved in both spermidine and spermine biosynthesis, leads to a bushy and dwarf phenotype in *Arabidopsis* by affecting cytokinin homeostasis (Cui et al., 2010). This mutant, namely *bud2-1*, has 25% higher levels of putrescine in comparison to the wild-type. Apparently, *bud2-1* has also enhanced root growth, supporting previous work that suggests putrescine as a growth promoter (Cui et al., 2010). Cytokinin levels are controlled by auxin (IAA) during bud outgrowth via apical dominance maintenance (Müller and Leyser, 2011). In this sense, opposite to putrescine, myo-inositol displays a positive correlation with sucrose. This glycoside conjugates IAA to temporarily control its availability, being hydrolyzed to set free IAA (Kowalczyk et al., 2003). Moreover, IAA conjugates with amino acids in plants, but only few conjugates (e.g., IAA-Ala, -Leu, and -Phe) are hydrolyzed to form free IAA. IAA-Asp and -Glu are in the degradation pathway or inhibition of the IAA action as IAA-Trp (Ludwig-Müller, 2011). Myo-inositol and galacturonate pathways are interconnected for ascorbate biosynthesis (Shen et al., 2009; Zhang et al., 2009), which is necessary for cell division and elongation (Tullio et al., 1999), biosynthesis of secondary metabolites and phytohormones (Smith et al., 2007). Taken together, our results suggested that the culm metabolism encompasses a complex metabolic network and confirmed the dual function fulfilled by this tissue: its initial sink role is replaced by the novel task as a nutrient source for the emergence of a new organ or seedling during the development of axillary meristem and bud outgrowth.

As the metabolic network of the culm unravel metabolites with putative role on bud outgrowth, we next investigated whether the metabolic composition among the selected commercial cultivars could be associated with sprouting rate. Our data shows that bud metabolome solely cannot explain the differences in the sprouting rate among the genotypes. In contrast, the culm metabolome could be used to classify at least the most contrasting genotypes. Interestingly, genotypes with higher sprouting rates tended to

have higher levels of certain metabolites, as it was the case of putrescine, whereas genotypes with low sprouting rates presented higher levels of sugars and amino acids, especially the branched-chain amino acids. These metabolites were positively correlated within their groups but were negatively correlated to each other in the metabolite-metabolite network. These findings suggest that carbon and nitrogen metabolism is not only involved on bud outgrowth but can also regulate this process mediating crosstalk with signaling pathways as, for example, forming conjugated compounds with phytohormones (Kowalczyk et al., 2003; Ludwig-Müller, 2011). Such approach has the potential for selecting metabolic markers and pathways associated to a certain agronomic traits in sugarcane as it was already successfully shown for other crop species (Meyer et al., 2007; Toubiana et al., 2012; Witt et al., 2012).

Our data also demonstrated that the sugarcane metabolome and bud sprouting rate are partially influenced by the genotype at least for the studied cultivars. As these commercial cultivars share part of their genetic background, we considered their genetic relatedness using pedigree information. In breeding programs, sugarcane flowering requires specific environmental conditions and it is highly genotype-dependent. Therefore, synchronization of panicles and flowering time is a challenge, and in many cases, precludes the accomplishment of desirable bi-parental crosses. In order to circumvent this limitation, multi-parental crosses can be performed as an alternative to achieve seed production. In those crosses, only the identity of the mother plant is known, and the pollens come freely from diverse male individuals. Out of the 16 genotypes used in this work, 7 presented an unknown male parental, indicating that these genotypes were obtained from multi-parental crosses. Furthermore, some genotypes, as TUC71-7, SP70-1143, RB855536 and particularly RB72454, are parental of several pedigree crosses leading to an overrepresentation of their genetic background in the selected genotypes, indicating the presence of genetic relatedness (kinship) in this panel.

Despite the incomplete record of both parentals in multi-crosses presented in this pedigree, it was partially possible to correlate bud sprouting and metabolome with the genetic information. We speculate that this correlation would be higher if more contrasting cultivars were analyzed. Even so, these results suggest that metabolic profile can be partially conserved at the parent-progeny degree in sugarcane, but not at more distant parentage levels. The use of metabolome as a proxy for genetics is appealing in sugarcane due to the complexity of its genome. Our results indicate that this approach should be feasible, opening the perspective for its application to assist sugarcane breeding programs.

## CONCLUSION

The work presented here clearly shows how metabolomics can be used to enrich the understanding of agronomic traits dependent on metabolic composition focusing on bud sprouting, a crucial process determining yield in sugarcane. Variability in metabolic features were identified even under a narrow genetic background typical for modern sugarcane cultivars.



Metabolite–metabolite correlation analysis was performed within and between tissues in order to add information on how the metabolism of buds and culms interact to promote sprouting. Metabolic networks revealed more complex patterns for culms in relation to buds, and enabled the recognition of key metabolites (e.g., sucrose, putrescine, glutamate, serine, and myo-inositol) affecting sprouting ability. Finally, those results were associated with the genetic background of each cultivar, showing that metabolites can be potentially used as indicators for the genetic background. Analysis of association panels with broader genetic variability and the use of informative genetic molecular markers could be used in the future to confirm the predictive power of metabolomics.

## AUTHOR CONTRIBUTIONS

DAF, MSC, and CC conceived the study. LGFA, JAA, and LDW performed the experiments. DAF, AC-G, RRA, and CC analyzed the data. MCMM, DAF, AC-G, MSC, and CC wrote the manuscript. MSC and HPH provided the supportive information.

## FUNDING

This work was supported by Max Planck Society and CNPq grant 402755/2012-0.

## ACKNOWLEDGMENTS

We thank Vinicius Fernandes de Souza for technical assistance and Isabella Valadão and Sandro Augusto Ferrarez for their support in the field experiments. We are also grateful to LabMET at CNPEM.

## REFERENCES

- Alex, R., Swanson, A., Humphrey, T., Chapman, R., McGarvey, B., Pocs, R., et al. (2004). Functional genomics uncovers three glucosyltransferases involved in the synthesis of the major sweet glucosides of *Stevia rebaudiana*. *Plant J.* 41, 56–67. doi: 10.1111/j.1365-313X.2004.02275.x
- Amadeu, R. R., Cellon, C., Olmstead, J. W., Garcia, A. A. F., Resende, M. F. R., and Muñoz, P. R. (2016). AGHmatrix: R package to construct relationship matrices for autotetraploid and diploid species: a blueberry example. *Plant Genome* 9, 1–10. doi: 10.3835/plantgenome2016.01.0009
- Baracat-Neto, J., Scarpore, F. V., De Araújo, R. B., and Scarpore-filho, J. A. (2017). Initial development and yield in sugarcane from different propagules. *Pesqui. Agropecu. Trop.* 47, 273–278.
- Barbier, F., Péron, T., Lecerf, M., Perez-Garcia, M. D., Barrière, Q., Rolčík, J., et al. (2015). Sucrose is an early modulator of the key hormonal mechanisms controlling bud outgrowth in *Rosa hybrida*. *J. Exp. Bot.* 66, 2569–2582. doi: 10.1093/jxb/erv047
- Barbier, F. F., Dun, E. A., and Beveridge, C. A. (2017). Apical dominance. *Curr. Biol.* 27, R864–R865. doi: 10.1016/j.cub.2017.05.024
- Barbier, F. F., Lunn, J. E., and Beveridge, C. A. (2015). Ready, steady, go! A sugar hit starts the race to shoot branching. *Curr. Opin. Plant Biol.* 25, 39–45. doi: 10.1016/j.pbi.2015.04.004
- Bernillon, S., Biais, B., Deborde, C., Maucourt, M., Cabasson, C., Gibon, Y., et al. (2013). Metabolomic and elemental profiling of melon fruit quality as affected

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.00857/full#supplementary-material>

**FIGURE S1 |** Pedigree of the sixteen selected sugarcane genotypes and their corresponding parentals. Gray circles represent parental genotypes that were not evaluated in this study; red, orange, blue, and green circles are genotypes ranked as low, intermediate-low, intermediate-high, and high sprouting, respectively. The arrows connecting genotypes are in the same color as their respective relatedness.

**FIGURE S2 |** Numerator relationship matrix among selected sugarcane commercial cultivars. The colors indicate the grade of relationship from low (red) to high (blue).

**FIGURE S3 |** Box plots of sprouting index of the sixteen selected sugarcane genotypes. For comparison among cultivars, the sprouting average considering the quartile analysis to classify them as low, intermediate-low, intermediate-high, and high sprouting was plotted. The groups are displayed in red, orange, blue, and green, respectively.

**FIGURE S4 |** Effect of genotypes on the levels of individual metabolites. Histograms show the number of metabolites whose levels changed according to the significance indicated by *P*-values. Bonferroni-corrected ANOVA was used to evaluate the effects of genotypes on culm (A) and bud (B).

**TABLE S1 |** Summary of the selected sugarcane genotypes and classification of their agronomic traits. Genotypes are coded from 1 to 16 and ranked as low (red), median-low (orange), intermediate-high (blue), and high sprouting (green) ability. \*POL: percentage by weight of apparent sucrose; <sup>1</sup>Data obtained from RIDESA breeding program (Carneiro, personal communication). <sup>2</sup>Data obtained from greenhouse experiments performed in this study.

**TABLE S2 |** Levels of all detected metabolites in culm and bud of the selected sugarcane cultivars.

**TABLE S3 |** Metabolite–metabolite correlations for bud, culm, and common to both tissues. Significant pairwise correlations within and between tissues ( $r \geq 0.5$ ,  $p \leq 0.05$ ) were highlighted in blue and yellow, representing positive and negative correlations, respectively.

- by genotype and environment. *Metabolomics* 9, 57–77. doi: 10.1007/s11306-012-0429-1
- Botha, F. C., and Black, K. G. (2000). Sucrose phosphate synthase and sucrose synthase activity during maturation of internodal tissue in sugarcane. *Aust. J. Plant Physiol.* 27, 81–85. doi: 10.1071/TT99098
- Botha, F. C., Lakshman, P., O'Connell, A., and Moore, P. H. (2013). "Hormones and growth regulators," in *Sugarcane: Physiology, Biochemistry, and Functional Biology*, eds P. H. Moore and F. C. Botha (Hoboken, NJ: Wiley). doi: 10.1002/9781118771280.ch14
- Boussiengui-Boussiengui, G., Groenewald, J. H., and Botha, F. C. (2016). Metabolic changes associated with the sink-source transition during sprouting of the axillary buds on the sugarcane culm. *Trop. Plant Biol.* 9, 1–11. doi: 10.1007/s12042-015-9158-8
- Brewer, P. B., Dun, E. A., Gui, R., Mason, M. G., and Beveridge, C. A. (2015). Strigolactone inhibition of branching independent of polar auxin transport. *Plant Physiol.* 168, 1820–1829. doi: 10.1104/pp.15.00014
- Cargnin, A., Müller, J. A., Mello, F. D. A., and Fogaça, C. M. (2008). *Brotação de Variedades de Cana-Deaçúcar Nas Condições de Cerrado Do Brasil-Central*. Available at: [http://simposio.cpac.embrapa.br/simposio/trabalhos\\_pdf/00526\\_trab1\\_ap.pdf](http://simposio.cpac.embrapa.br/simposio/trabalhos_pdf/00526_trab1_ap.pdf)
- Carneiro, A. E. V., Trivelin, P. C. O., and Victoria, R. L. (1995). Utilização da reserva orgânica e de nitrogênio do tolete de plantio (colmo-semente) no desenvolvimento da cana-planta. *Sci. Agric.* 52, 199–209. doi: 10.1590/S0103-90161995000200001

- Cascales-Minana, B., Munoz-Bertomeu, J., Flores-Tornero, M., Anoman, A. D., Pertusa, J., Alaiz, M., et al. (2013). The phosphorylated pathway of serine biosynthesis is essential both for male gametophyte and embryo development and for root growth in *Arabidopsis*. *Plant Cell* 25, 2084–2101. doi: 10.1105/tpc.113.112359
- Chellamuthu, V. R., Ermilova, E., Lapina, T., Lüddecke, J., Minaeva, E., Herrmann, C., et al. (2014). A widespread glutamine-sensing mechanism in the plant kingdom. *Cell* 159, 1188–1199. doi: 10.1016/j.cell.2014.10.015
- Cheng, J., Galili, T., Bostock, M., and Palmer, J. (2018). *Interactive Heat Maps Using 'htmlwidgets' and 'D3.js.'* Available at: <https://cran.r-project.org/web/packages/d3heatmap/d3heatmap.pdf>
- Cuadros-Inostroza, A., Caldana, C., Redestig, H., Kusano, M., Lisec, J., Peña-Cortés, H., et al. (2009). TargetSearch - a bioconductor package for the efficient preprocessing of GC-MS metabolite profiling data. *BMC Bioinformatics* 10:428. doi: 10.1186/1471-2105-10-428
- Cui, X., Ge, C., Wang, R., Wang, H., Chen, W., Fu, Z., et al. (2010). The BUD2 mutation affects plant architecture through altering cytokinin and auxin responses in *Arabidopsis*. *Cell Res.* 20, 576–586. doi: 10.1038/cr.2010.51
- Dan, Z., Hu, J., Zhou, W., Yao, G., Zhu, R., Zhu, Y., et al. (2016). Metabolic prediction of important agronomic traits in hybrid rice (*Oryza sativa* L.). *Sci. Rep.* 6:21732. doi: 10.1038/srep21732
- de Setta, N., Monteiro-Vitorello, C. B., Metcalfe, C. J., Cruz, G. M. Q., Bem, L. E. D., Vicentini, R., et al. (2014). Building the sugarcane genome for biotechnology and identifying evolutionary trends. *BMC Genomics* 15:540. doi: 10.1186/1471-2164-15-540
- Dingkuhn, M., Luquet, D., Tambour, L., Kim, H. K., and Song, Y. H. (2007). “Is plant growth driven by sink regulation?” in *Scale and Complexity in Plant Systems Research: Gene-Plant-Crop Relations*, eds P. C. Struik and H. H. Van Laar (Berlin: Springer), 157–170.
- Domagalska, M. A., and Leyser, O. (2011). Signal integration in the control of shoot branching. *Nat. Rev. Mol. Cell Biol.* 12, 211–221. doi: 10.1038/nrm3088
- Dun, E. A., de Saint-Germain, A., Rameau, C., and Beveridge, C. A. (2012). Antagonistic action of strigolactone and cytokinin in bud outgrowth control. *Plant Physiol.* 158, 487–498. doi: 10.1104/pp.111.186783
- Durbak, A., Yao, H., and McSteen, P. (2012). Hormone signaling in plant development. *Curr. Opin. Plant Biol.* 15, 92–96. doi: 10.1016/j.pbi.2011.12.004
- Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D. M., and Borsboom, D. (2012). Qgraph: network visualizations of relationships in psychometric data. *J. Stat. Softw.* 48, 1–18. doi: 10.18637/jss.v048.i04
- Fernie, A. R., and Tohge, T. (2017). The genetics of plant metabolism. *Annu. Rev. Genet.* 51, 287–310. doi: 10.1146/annurev-genet-120116-024640
- Fichtner, F., Barbier, F. F., Feil, R., Watanabe, M., Annunziata, M. G., Chabikwa, T. G., et al. (2017). Trehalose 6-phosphate is involved in triggering axillary bud outgrowth in garden pea (*Pisum sativum* L.). *Plant J.* 92, 611–623. doi: 10.1111/tpj.13705
- Galili, G., and Höfgen, R. (2002). Metabolic engineering of amino acids and storage proteins in plants. *Metab. Eng.* 4, 3–11. doi: 10.1006/mben.2001.0203
- Ge, C., Cui, X., Wang, Y., Hu, Y., Fu, Z., Zhang, D., et al. (2006). BUD2, encoding an S-adenosylmethionine decarboxylase, is required for arabidopsis growth and development. *Cell Res.* 16, 446–456. doi: 10.1038/sj.cr.7310056
- Giavalisco, P., Li, Y., Matthes, A., Eckhardt, A., Hubberten, H. M., Hesse, H., et al. (2011). Elemental formula annotation of polar and lipophilic metabolites using <sup>13</sup>C, <sup>15</sup>N and <sup>34</sup>S isotope labelling, in combination with high-resolution mass spectrometry. *Plant J.* 68, 364–376. doi: 10.1111/j.1365-313X.2011.04682.x
- Gill, S. S., and Tuteja, N. (2010). Polyamines and abiotic stress tolerance in plants. *Plant Signal. Behav.* 5, 26–33. doi: 10.4161/psb.5.1.10291
- Glaubitx, U., Li, X., Köhl, K. I., van-Dongen, J. T., Hincha, D. K., and Zuther, E. (2014). Differential physiological responses of different rice (*Oryza sativa*) cultivars to elevated night temperature during vegetative growth. *Funct. Plant Biol.* 41, 437–448. doi: 10.1071/FP13132
- Goulet, C., Kamiyoshihara, Y., Lam, N. B., Richard, T., Taylor, M. G., Tieman, D. M., et al. (2015). Divergence in the enzymatic activities of a tomato and *Solanum pennellii* alcohol acyltransferase impacts fruit volatile ester composition. *Mol. Plant* 8, 153–162. doi: 10.1016/j.molp.2014.11.007
- Hoang, N. V., Furtado, A., Mason, P. J., Marquardt, A., Kasirajan, L., Thirugnanasambandam, 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
- Huang, M., Abel, C., Sohrabi, R., Petri, J., Haupt, I., Cosimano, J., et al. (2010). Variation of herbivore-induced volatile terpenes among *Arabidopsis* ecotypes depends on allelic differences and subcellular targeting of two terpene synthases, TPS02 and TPS03. *Plant Physiol.* 153, 1293–1310. doi: 10.1104/pp.110.154864
- Huang, X., Effgen, S., Meyer, R. C., Theres, K., and Koornneef, M. (2012). Epistatic natural allelic variation reveals a function of AGAMOUS-LIKE6 in axillary bud formation in *Arabidopsis*. *Plant Cell* 24, 2364–2379. doi: 10.1105/tpc.112.099168
- Jaillais, Y., and Chory, J. (2010). Unraveling the paradoxes of plant hormone signaling integration. *Nat. Struct. Mol. Biol.* 17, 642–645. doi: 10.1038/nsmb0610-642
- Jain, R., Solomon, S., Shrivastava, A. K., and Lal, P. (2009). Nutrient application improves stubble bud sprouting under low temperature conditions in sugarcane. *Sugar Tech* 11, 83–85. doi: 10.1007/s12355-009-0016-6
- Kebrom, T. H., and Mullet, J. E. (2015). Photosynthetic leaf area modulates tiller bud outgrowth in sorghum. *Plant Cell Environ.* 38, 1471–1478. doi: 10.1111/pce.12500
- Kooke, R., and Keurentjes, J. J. B. (2012). Multi-dimensional regulation of metabolic networks shaping plant development and performance. *J. Exp. Bot.* 63, 3353–3365. doi: 10.1093/jxb/err373
- Kowalczyk, S., Jakubowska, A., Zielinska, E., and Bandurski, R. S. (2003). Bifunctional indole-3-acetyl transferase catalyses synthesis and hydrolysis of indole-3-acetyl-myoinositol in immature endosperm of *Zea mays*. *Physiol. Plant.* 119, 165–174. doi: 10.1034/j.1399-3054.2003.00158.x
- Kumar, R., Bohra, A., Pandey, A. K., Pandey, M. K., and Kumar, A. (2017). Metabolomics for plant improvement: status and prospects. *Front. Plant Sci.* 8:1302. doi: 10.3389/fpls.2017.01302
- Kusano, M., Tabuchi, M., Fukushima, A., Funayama, K., Diaz, C., Kobayashi, M., et al. (2011). Metabolomics data reveal a crucial role of cytosolic glutamine synthetase 1;1 in coordinating metabolic balance in rice. *Plant J.* 66, 456–466. doi: 10.1111/j.1365-313X.2011.04506.x
- Kusano, T., Yamaguchi, K., Berberich, T., and Takahashi, Y. (2007). Advances in polyamine research in 2007. *J. Plant Res.* 120, 345–350. doi: 10.1007/s10265-007-0074-3
- Lea, P. J., and Mifflin, B. J. (1974). Alternative route for nitrogen assimilation in higher plants. *Nature* 251, 614–616. doi: 10.1038/251614a0
- Lisec, J., Meyer, R. C., Steinfath, M., Redestig, H., Becher, M., Witucka-Wall, H., et al. (2008). Identification of metabolic and biomass QTL in *Arabidopsis thaliana* in a parallel analysis of RIL and IL populations. *Plant J.* 53, 960–972. doi: 10.1111/j.1365-313X.2007.03383.x
- Ludwig-Müller, J. (2011). Auxin conjugates: their role for plant development and in the evolution of land plants. *J. Exp. Bot.* 62, 1757–1773. doi: 10.1093/jxb/erq412
- Lunn, J. E. (2007). Compartmentation in plant metabolism. *J. Exp. Bot.* 58, 35–47. doi: 10.1093/jxb/erl134
- Manhães, C. M. C., Garcia, R. F., Francelino, F. M. A., Francelino, H. O., and Coelho, F. C. (2015). Factors that affect sprouting and tillering of sugar cane. *Rev. Vertices* 17, 163–181. doi: 10.5935/1809-2667.20150011
- Mason, M. G., Ross, J. J., Babst, B. A., Wienclaw, B. N., and Beveridge, C. A. (2014). Sugar demand, not auxin, is the initial regulator of apical dominance. *Proc. Natl. Acad. Sci. U.S.A.* 111, 6092–6097. doi: 10.1073/pnas.1322045111
- Meyer, R. C., Steinfath, M., Lisec, J., Becher, M., Witucka-Wall, H., Törjék, O., et al. (2007). The metabolic signature related to high plant growth rate in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 104, 4759–4764. doi: 10.1073/pnas.0609709104
- Monti, L. L., Bustamante, C. A., Osorio, S., Gabilondo, J., Borsani, J., Lauxmann, M. A., et al. (2016). Metabolic profiling of a range of peach fruit varieties reveals high metabolic diversity and commonalities and differences during ripening. *Food Chem.* 190, 879–888. doi: 10.1016/j.foodchem.2015.06.043
- Müller, D., and Leyser, O. (2011). Auxin, cytokinin and the control of shoot branching. *Ann. Bot.* 107, 1203–1212. doi: 10.1093/aob/mcr069
- Nunes-Nesi, A., Fernie, A. R., and Stitt, M. (2010). Metabolic and signaling aspects underpinning the regulation of plant carbon nitrogen interactions. *Mol. Plant* 3, 973–996. doi: 10.1093/mp/ssq049
- Obata, T., Witt, S., Lisec, J., Palacios-Rojas, N., Florez-Sarasa, I., Yousfi, S., et al. (2015). Metabolite profiles of maize leaves in drought, heat, and combined stress

- field trials reveal the relationship between metabolism and grain yield. *Plant Physiol.* 169, 2665–2683. doi: 10.1104/pp.15.01164
- Ohashi, M., Ishiyama, K., Kusano, M., Fukushima, A., Kojima, S., Hanada, A., et al. (2015). Lack of cytosolic glutamine synthetase1;2 in vascular tissues of axillary buds causes severe reduction in their outgrowth and disorder of metabolic balance in rice seedlings. *Plant J.* 81, 347–356. doi: 10.1111/tpj.12731
- Oliveira, J. C. M., Timm, L. C., Tominaga, T. T., Cássaro, F. A. M., Reichardt, K., Bacchi, O. O. S., et al. (2001). Soil temperature in a sugar-cane crop as a function of the management system. *Plant Soil* 230, 61–66. doi: 10.1023/A:1004820119399
- O'Neill, B. P., Purnell, M. P., Anderson, D. J., Nielsen, L. K., and Brumbley, S. M. (2012). Sucrose mobilisation in sugarcane stalk induced by heterotrophic axillary bud growth. *Trop. Plant Biol.* 5, 173–182. doi: 10.1007/s12042-012-9097-6
- Patrick, J. W., and Colyvas, K. (2014). Crop yield components – photoassimilate supply- or utilisation limited-organ development? *Funct. Plant Biol.* 41, 893–913. doi: 10.1071/FP14048
- Paul, M. J., and Foyer, C. H. (2001). Sink regulation of photosynthesis. *J. Exp. Bot.* 52, 1383–1400. doi: 10.1093/jxb/52.360.1383
- Pereira, L. F. M., Ferreira, V. M., de Oliveira, N. G., Sarmento, P. L. V. S., Endres, L., and Teodoro, I. (2017). Sugars levels of four sugarcane genotypes in different stem portions during the maturation phase. *An. Acad. Bras. Cienc.* 89, 1231–1242. doi: 10.1590/0001-3765201720160594
- Purohit, S. D., Singhvi, A., Nagori, R., and Vyas, S. (2007). Polyamines stimulate shoot bud proliferation and multiplication in *Achras sapota* grown in culture. *Indian J. Biotechnol.* 6, 85–90.
- Qin, X., Coku, A., Inoue, K., and Tian, L. (2011). Expression, subcellular localization, and cis-regulatory structure of duplicated phytoene synthase genes in melon (*Cucumis melo* L.). *Planta* 234, 737–748. doi: 10.1007/s00425-011-1442-8
- Rameau, C., Bertheloot, J., Leduc, N., Andrieu, B., Foucher, F., and Sakr, S. (2015). Multiple pathways regulate shoot branching. *Front. Plant Sci.* 5:741. doi: 10.3389/fpls.2014.00741
- Ren, L. L., Liu, Y. J., Liu, H. J., Qian, T. T., Qi, L. W., Wang, X. R., et al. (2014). Subcellular relocalization and positive selection play key roles in the retention of duplicate genes of *Populus* class III peroxidase family. *Plant Cell* 26, 2404–2419. doi: 10.1105/tpc.114.124750
- Riaño-Pachón, D. M., and Mattiello, L. (2017). Draft genome sequencing of the sugarcane hybrid SP80-3280. *F1000Res.* 6:861. doi: 10.12688/f1000research.11859.2
- Roessner, U., Luedemann, A., Brust, D., Fiehn, O., Linke, T., Willmitzer, L., et al. (2001). Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell* 13, 11–29. doi: 10.1105/tpc.13.1.11
- Rohart, F., Gautier, B., Singh, A., and Cao, K. A. L. (2017). MixOmics: an R package for 'omics feature selection and multiple data integration. *PLoS Comput. Biol.* 13:e1005752. doi: 10.1371/journal.pcbi.1005752
- Ros, R., Muñoz-Bertomeu, J., and Krueger, S. (2014). Serine in plants: biosynthesis, metabolism, and functions. *Trends Plant Sci.* 19, 564–569. doi: 10.1016/j.tplants.2014.06.003
- Rose, S., and Botha, F. C. (2000). Distribution patterns of neutral invertase and sugar content in sugarcane internodal tissues. *Plant Physiol. Biochem.* 38, 819–824. doi: 10.1016/S0981-9428(00)01190-6
- Saito, K., and Matsuda, F. (2010). Metabolomics for functional genomics, systems biology, and biotechnology. *Annu. Rev. Plant Biol.* 61, 463–489. doi: 10.1146/annurev-arplant.043008.092035
- Schauer, N., Semel, Y., Roessner, U., Gur, A., Balbo, I., Carrari, F., et al. (2006). Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nat. Biotechnol.* 24, 447–454. doi: 10.1038/nbt1192
- Seki, H., Sawai, S., Ohshima, K., Mizutani, M., Ohnishi, T., Sudo, H., et al. (2011). Triterpene functional genomics in licorice for identification of CYP72A154 involved in the biosynthesis of glycyrrhizin. *Plant Cell* 23, 4112–4123. doi: 10.1105/tpc.110.082685
- Shen, C. H., Krishnamurthy, R., and Yeh, K. W. (2009). Decreased L-ascorbate content mediating bolting is mainly regulated by the galacturonate pathway in *Oncidium*. *Plant Cell Physiol.* 50, 935–946. doi: 10.1093/pcp/pcp045
- Singels, A., and Smit, M. A. (2002). The effect of row spacing on an irrigated plant crop of sugarcane variety Nco376. *Proc. South Afr. Sugar Technol. Assoc.* 76, 94–105.
- Singh, R. K., Verma, H. K., Singh, S. P., Rastogi, J., and Sharma, B. L. (2016). Study on morpho-physiological differences in sugarcane cultivars varying in germination efficiency. *Braz. J. Bot.* 39, 123–130. doi: 10.1007/s40415-015-0224-0
- Smeekeens, S., Ma, J., Hanson, J., and Rolland, F. (2010). Sugar signals and molecular networks controlling plant growth. *Curr. Opin. Plant Biol.* 13, 273–278. doi: 10.1016/j.pbi.2009.12.002
- Smit, M. A. (2011). Characterising the factors that affect germination and emergence in sugarcane. *Int. Sugar J.* 113, 65–67.
- Smith, A. G., Croft, M. T., Moulin, M., and Webb, M. E. (2007). Plants need their vitamins too. *Curr. Opin. Plant Biol.* 10, 266–275. doi: 10.1016/j.pbi.2007.04.009
- Smith, A. M., and Stitt, M. (2007). Coordination of carbon supply and plant growth. *Plant Cell Environ.* 30, 1126–1149. doi: 10.1111/j.1365-3040.2007.01708.x
- Soetaert, K. (2017). Package 'Diagram: Functions for Visualising Simple Graphs (Networks), Plotting Flow Diagrams'. CRAN. Available at: <https://cran.r-project.org/web/packages/diagram/diagram.pdf>
- Soltis, N. E., and Kliebenstein, D. J. (2015). Natural variation of plant metabolism: genetic mechanisms, interpretive caveats, and evolutionary and mechanistic insights. *Plant Physiol.* 169, 1456–1468. doi: 10.1104/pp.15.01108
- Song, J., Yang, X., Resende, M. F. R., Neves, L. G., Todd, J., Zhang, J., et al. (2016). Natural allelic variations in highly polyploidy *Saccharum* complex. *Front. Plant Sci.* 7:804. doi: 10.3389/fpls.2016.00804
- Sprenger, H., Erban, A., Seddig, S., Rudack, K., Thalhammer, A., Le, M. Q., et al. (2018). Metabolite and transcript markers for the prediction of potato drought tolerance. *Plant Biotechnol. J.* 16, 939–950. doi: 10.1111/pbi.12840
- Stacklies, W., Redestig, H., Scholz, M., Walther, D., and Selbig, J. (2007). PcaMethods—a bioconductor package providing PCA methods for incomplete data. *Bioinformatics* 23, 1164–1167. doi: 10.1093/bioinformatics/btm069
- Stitt, M., Sulpice, R., and Keurentjes, J. (2010). Metabolic networks: how to identify key components in the regulation of metabolism and growth. *Plant Physiol.* 152, 428–444. doi: 10.1104/pp.109.150821
- Sweetlove, L. J., and Fernie, A. R. (2013). The spatial organization of metabolism within the plant cell. *Annu. Rev. Plant Biol.* 64, 723–746. doi: 10.1146/annurev-arplant-050312-120233
- Tarancon, C., González-Grandío, E., Oliveros, J. C., Nicolas, M., and Cubas, P. (2017). A conserved carbon starvation response underlies bud dormancy in woody and Herbaceous species. *Front. Plant Sci.* 8:788. doi: 10.3389/fpls.2017.00788
- Todaka, D., Zhao, Y., Yoshida, T., Kudo, M., Kidokoro, S., Mizoi, J., et al. (2017). Temporal and spatial changes in gene expression, metabolite accumulation and phytohormone content in rice seedlings grown under drought stress conditions. *Plant J.* 90, 61–78. doi: 10.1111/tpj.13468
- Tohge, T., Nishiyama, Y., Hirai, M. Y., Yano, M., Nakajima, J. I., Awazuhara, M., et al. (2005). Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing an MYB transcription factor. *Plant J.* 42, 218–235. doi: 10.1111/j.1365-313X.2005.02371.x
- Toubiana, D., Batushansky, A., Tzfadia, O., Scossa, F., Khan, A., Barak, S., et al. (2015). Combined correlation-based network and MQTL analyses efficiently identified loci for branched-chain amino acid, serine to threonine, and proline metabolism in tomato seeds. *Plant J.* 81, 121–133. doi: 10.1111/tpj.12717
- Toubiana, D., Semel, Y., Tohge, T., Beleggia, R., Cattivelli, L., Rosental, L., et al. (2012). Metabolic profiling of a mapping population exposes new insights in the regulation of seed metabolism and seed, fruit, and plant relations. *PLoS Genet.* 8:e1002612. doi: 10.1371/journal.pgen.1002612
- Tullio, M. C. D., Paciolla, C., Vecchia, F. D., Rascio, N., Emerico, S. D., Gara, L. D., et al. (1999). Changes in onion root development induced by the inhibition of peptidyl-prolyl hydroxylase and influence of the ascorbat on cell division and elongation. *Planta* 209, 424–434. doi: 10.1007/s004250050745
- Umebara, M., Hanada, A., Yoshida, S., Akiyama, K., Arite, T., Takeda-Kamiya, N., et al. (2008). Inhibition of shoot branching by new terpenoid plant hormones. *Nature* 455, 195–200. doi: 10.1038/nature07272
- Urano, K., Hobo, T., and Shinozaki, K. (2005). *Arabidopsis* ADC genes involved in polyamine biosynthesis are essential for seed development. *FEBS Lett.* 579, 1557–1564. doi: 10.1016/j.febslet.2005.01.048

- Uys, L., Botha, F. C., Hofmeyr, J. H. S., and Rohwer, J. M. (2007). Kinetic model of sucrose accumulation in maturing sugarcane culm tissue. *Phytochemistry* 68, 2375–2392. doi: 10.1016/j.phytochem.2007.04.023
- Verma, A. K., Agarwal, A. K., Dubey, R. S., Solomon, S., and Singh, S. B. (2013). Sugar partitioning in sprouting lateral bud and shoot development of sugarcane. *Plant Physiol. Biochem.* 62, 111–115. doi: 10.1016/j.plaphy.2012.10.021
- Vorster, D. J., and Botha, F. C. (1999). Sugarcane internodal invertases and tissue maturity. *J. Plant Physiol.* 155, 470–476. doi: 10.1016/S0176-1617(99)80041-8
- Weckwerth, W., Wenzel, K., and Fiehn, O. (2004). Process for the integrated extraction, identification and quantification of metabolites, proteins and RNA to reveal their co-regulation in biochemical networks. *Proteomics* 4, 78–83. doi: 10.1002/pmic.200200500
- Wei, T., and Simko, V. (2017). *R Package Corrplot: Visualization of a Correlation Matrix (Version 0.84)*. Available at: <https://github.com/taiyun/corrplot>
- Wen, W., Li, K., Alseekh, S., Omranian, N., Zhao, L., Zhou, Y., et al. (2015). Genetic determinants of the network of primary metabolism and their relationships to plant performance in a maize recombinant inbred line population. *Plant Cell* 27, 1839–1856. doi: 10.1105/tpc.15.00208
- Whittaker, A., and Botha, F. C. (1997). Carbon partitioning during sucrose accumulation in sugarcane internodal tissue. *Plant Physiol.* 115, 1651–1659. doi: 10.1104/pp.115.4.1651
- Wimmer, V., Albrecht, T., Auinger, H. J., and Schön, C. C. (2012). Synbreed: a framework for the analysis of genomic prediction data using R. *Bioinformatics* 28, 2086–2087. doi: 10.1093/bioinformatics/bts335
- Witt, S., Galicia, L., Lise, J., Cairns, J., Tiessen, A., Araus, J. L., et al. (2012). Metabolic and phenotypic responses of greenhouse-grown maize hybrids to experimentally controlled drought stress. *Mol. Plant* 5, 401–417. doi: 10.1093/mp/ssf102
- Wright, S. (1922). Coefficients of inbreeding and relationship. *Am. Nat.* 56, 330–338. doi: 10.1086/279872
- Xiong, Y., McCormack, M., Li, L., Hall, Q., Xiang, C., and Sheen, J. (2013). Glucose–TOR signalling reprograms the transcriptome and activates meristems. *Nature* 496, 181–186. doi: 10.1038/nature12030
- Yadav, U. P., Ivakov, A., Feil, R., Duan, G. Y., Walther, D., Giavalisco, P., et al. (2014). The sucrose–trehalose 6-Phosphate (Tre6P) nexus: specificity and mechanisms of sucrose signalling by Tre6P. *J. Exp. Bot.* 65, 1051–1068. doi: 10.1093/jxb/ert457
- Yamaya, T., and Oaks, A. (2004). “Metabolic regulation of ammonium uptake and assimilation,” in *Nitrogen Acquisition and Assimilation in Higher Plants*, eds S. Amâncio and I. Stulen (Dordrecht: Springer), 35–63. doi: 10.1007/978-1-4020-2728-4\_2
- Zhang, J., Nagai, C., Yu, Q., Pan, Y. B., Ayala-Silva, T., Schnell, R. J., et al. (2012). Genome size variation in three *Saccharum* species. *Euphytica* 185, 511–519. doi: 10.1007/s10681-012-0664-6
- Zhang, W., Lorence, A., Gruszewski, H. A., Chevone, B. I., and Nessler, C. L. (2009). *AMRI*, an Arabidopsis gene that coordinately and negatively regulates the mannose/L-galactose ascorbic acid biosynthetic pathway. *Plant Physiol.* 150, 942–950. doi: 10.1104/pp.109.138453
- Zheng, H., Rowland, O., and Kunst, L. (2005). Disruptions of the Arabidopsis enoyl-CoA reductase gene reveal an essential role for very-long-chain fatty acid synthesis in cell expansion during plant morphogenesis. *Plant Cell* 17, 1467–1481. doi: 10.1105/tpc.104.030155
- Zhu, Y. J., Komor, E., and Moore, P. H. (1997). Sucrose accumulation in the sugarcane stem is regulated by the difference between the activities of soluble acid invertase and sucrose phosphate synthase. *Plant Physiol.* 115, 609–616. doi: 10.1104/pp.115.2.609

**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.

The reviewer AF declared a shared affiliation, though no other collaboration, with one of the authors RRA to the handling Editor.

Copyright © 2018 Ferreira, Martins, Cheavegatti-Gianotto, Carneiro, Amadeu, Aricetti, Wolf, Hoffmann, de Abreu and Caldana. 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.





# Combined Drought and Heat Activates Protective Responses in *Eucalyptus globulus* That Are Not Activated When Subjected to Drought or Heat Stress Alone

Barbara Correia<sup>1†</sup>, Robert D. Hancock<sup>2†</sup>, Joana Amaral<sup>1</sup>, Aurelio Gomez-Cadenas<sup>3</sup>, Luis Valledor<sup>4</sup> and Glória Pinto<sup>1\*</sup>

<sup>1</sup> Department of Biology, Centre for Environmental and Marine Studies, University of Aveiro, Aveiro, Portugal, <sup>2</sup> Cell and Molecular Sciences, The James Hutton Institute, Dundee, United Kingdom, <sup>3</sup> Departamento de Ciencias Agrarias y del Medio Natural, Universitat Jaume I, Castellón de la Plana, Spain, <sup>4</sup> Department of Organisms and Systems Biology, University of Oviedo, Oviedo, Spain

## OPEN ACCESS

### Edited by:

Atsushi Fukushima,  
RIKEN, Japan

### Reviewed by:

Ernani Pinto,  
Universidade de São Paulo, Brazil  
Carlos Alberto Labate,  
Universidade de São Paulo, Brazil  
Kris Morreel,  
Flemish Institute for Technological  
Research, Belgium

### \*Correspondence:

Glória Pinto  
gpinto@ua.pt

<sup>†</sup> These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Plant Systems and Synthetic Biology,  
a section of the journal  
Frontiers in Plant Science

**Received:** 17 October 2017

**Accepted:** 28 May 2018

**Published:** 20 June 2018

### Citation:

Correia B, Hancock RD, Amaral J,  
Gomez-Cadenas A, Valledor L and  
Pinto G (2018) Combined Drought  
and Heat Activates Protective  
Responses in *Eucalyptus globulus*  
That Are Not Activated When  
Subjected to Drought or Heat Stress  
Alone. *Front. Plant Sci.* 9:819.  
doi: 10.3389/fpls.2018.00819

Aiming to mimic a more realistic field condition and to determine convergent and divergent responses of individual stresses in relation to their combination, we explored physiological, biochemical, and metabolomic alterations after drought and heat stress imposition (alone and combined) and recovery, using a drought-tolerant *Eucalyptus globulus* clone. When plants were exposed to drought alone, the main responses included reduced pre-dawn water potential ( $\Psi_{pd}$ ) and gas exchange. This was accompanied by increases in malondialdehyde (MDA) and total glutathione, indicative of oxidative stress. Abscissic acid (ABA) levels increased while the content of jasmonic acid (JA) fell. Metabolic alterations included reductions in the levels of sugar phosphates accompanied by increases in starch and non-structural carbohydrates. Levels of  $\alpha$ -glycerophosphate and shikimate were also reduced while free amino acids increased. On the other hand, heat alone triggered an increase in relative water content (RWC) and  $\Psi_{pd}$ . Photosynthetic rate and pigments were reduced accompanied by a reduction in water use efficiency. Heat-induced a reduction of salicylic acid (SA) and JA content. Sugar alcohols and several amino acids were enhanced by the heat treatment while starch, fructose-6-phosphate, glucose-6-phosphate, and  $\alpha$ -glycerophosphate were reduced. Contrary to what was observed under drought, heat stress activated the shikimic acid pathway. Drought-stressed plants subject to a heat shock exhibited a sharp decrease in gas exchange,  $\Psi_{pd}$  and JA, no alterations in electrolyte leakage, MDA, starch, and pigments and increased glutathione pool in relation to control. Comparing this with drought stress alone, subjecting drought stressed plants to an additional heat stress alleviated  $\Psi_{pd}$  and MDA, maintained an increased glutathione pool and reduced starch content and non-structural carbohydrates. A novel response triggered by the combined stress was the accumulation of cinnamate. Regarding recovery, most of the parameters affected by each stress condition reversed after re-establishment of control growing conditions. These results highlight that the combination of drought and

heat provides significant protection from more detrimental effects of drought-stressed eucalypts, confirming that combined stress alter plant metabolism in a novel manner that cannot be extrapolated by the sum of the different stresses applied individually.

**Keywords:** plant metabolism, isolated stress, combined stress, recovery, network analysis

## INTRODUCTION

Forest trees, as all sessile plants, have evolved many mechanisms that enable them to thrive in variable environmental conditions, ranging from circadian regulation (Dodd et al., 2005) to recovery from overpowering stress (Brodribb and Cochard, 2009). Despite these physiological adaptations, the long life-span of trees does not allow for rapid genetic adaptation to environmental changes, rendering forests particularly susceptible to climate change (Lindner et al., 2010). Therefore, climate-driven forest vulnerability and tree die-off have become emerging concerns for forest sustainability worldwide (Anderegg et al., 2012; Allen et al., 2015).

Decades of research have significantly improved our understanding of how abiotic stresses that plants encounter in the field, such as drought and heat stress, affect plant development and growth (Rennenberg et al., 2006). However, predominant abiotic stress factors have been mostly tested individually and under controlled laboratory conditions (Mittler and Blumwald, 2010). In contrast, relatively little attention has been given to the combined effects of abiotic stresses, for example, in the field water deficit does not occur alone but associated with high temperature or high light (Chaves et al., 2002).

There is a growing body of evidence that the impacts of a combination of different stress factors on plant functioning traits do not necessarily lead to an additive response but rather to unique responses as a consequence of a synergistic or antagonistic effect of both stress factors (Bansal et al., 2013; Pandey et al., 2015). The high degree of complexity results from the fact that when two stresses co-occur, plant adaptation to the stress combination is governed by the interaction of the two stresses, controlled by different signaling pathways that may interact, inhibit one another or be prioritized differentially by the plant (Zandalinas et al., 2017).

A particular abiotic stress induces a plant response tailored to that specific environmental condition and, when encountering different combined stresses, a plant might actually require conflicting adjustments (Mittler and Blumwald, 2010). Under combined drought and heat stress, for example, plants have to act and balance stomatal responses between preventing water loss and cooling their leaves by transpiration, meaning that a proper defense response depends simultaneously on decreasing and increasing stomatal conductance (Mittler and Blumwald, 2010).

The previous example leaves no doubt on the challenging task of researching abiotic stress combination. Several studies have already researched this subject mainly focusing on the drought and heat combination (Valladares and Pearcy, 1997; Dobra et al., 2010; Silva et al., 2010; Arend et al., 2013; Duan et al., 2014). The results indicate a plethora of plant responses ranging

from stomatal and non-stomatal limitations to photosynthesis (Arend et al., 2013), photo-inhibition (Valladares and Pearcy, 1997), changes in key stress signaling components, such as reactive oxygen species (Silva et al., 2010) and plant hormones (Dobra et al., 2010), up to rapid mortality through loss of stem hydraulic conductivity (Duan et al., 2014). Furthermore, the conclusions are very species/experiment dependent: elevated temperature is beneficial when imposed alone but is detrimental when combined with drought (Arend et al., 2013); elevated temperature triggers rapid mortality through hydraulic failure, which is induced by drought (Duan et al., 2014); drought greatly disturbs photosystem II activity and oxidative metabolism, which are strongly stimulated by heat stress (Silva et al., 2010). Given the known impact of abiotic stress on the plant metabolome (Warren et al., 2011; Hochberg et al., 2013), we would also expect extensive research on this topic, but the available knowledge is limited (Obata et al., 2015).

Among forest plantations, *Eucalyptus* species play an increasingly essential role to guarantee the world's demand for wood products, and assessing the impact of drought and heat on such economically important plants is highly pertinent since both factors are considered the main drivers controlling vulnerability of *Eucalyptus* plantations (Booth, 2013). Our previous research which compared results from a controlled climate chamber experiment with field-grown *Eucalyptus globulus* Labill. corroborated that the knowledge acquired from imposing the stress individually to test stress-tolerant plants cannot be extrapolated to field-grown plants (Correia et al., 2018). This urged us to evaluate the impact of combined drought and heat stress in *E. globulus* plants, mimicking a more realistic field condition. Since assessing recovery may also be very informative and provide better insights of the severity of the combined stress than observations done at the stress imposition (Mitchell et al., 2013), we have also included a post stress period.

This study hence arises from the need to elucidate the major responses that take place in *E. globulus* under combined drought and heat stress. Aiming to determine convergent and divergent responses of the individual stress in relation to their combination, we explored physiological and biochemical alterations after stress imposition (alone and combined) and recovery using a drought-tolerant clone. An additional key goal was to get an extra dimension by identifying and integrating major metabolomic alterations.

## MATERIALS AND METHODS

### Plant Material and Experimental Design

Rooted cuttings of *E. globulus* (clone AL-18) were obtained from the breeding program of Altri Florestal SA (Portugal) and transplanted to 1 L plastic pots filled with equal weight of a 3:2

(w/w) peat:perlite mixture. The potted cuttings were then divided and placed in two climate chambers (Fitoclima 1200, Aralab, Portugal) for a one-month acclimation period. Conditions were 25/20°C (day/night), 16/8 h (day/night) photoperiod, 50% relative humidity and  $600 \mu\text{mol m}^{-2} \text{s}^{-1}$  photosynthetic photon flux density. During the acclimation period, plants were watered up to 70% field capacity (FC) and fertilized weekly with a NPK (5:8:10) nutritive solution. Pot weight was monitored every day and the percentages of FC were maintained by adding the amount of water lost. During the experiment, environmental conditions inside the climate chambers were maintained as in the acclimation period and only watering was altered. Half of the cuttings in each climate chamber was assigned to a control well-watered regime (C: water supplied every day until soil water content reached 70% FC) and the other half was assigned to a drought regime (D: water supplied every day until soil content reached 18% FC). This lasted for 5 days. Air temperature inside the second climate chamber was then gradually increased, and plants from both groups (C and D) were subject to 40°C during 4 h (H – heat stress; and D\*H – combined (combination of both drought and heat) stress, respectively). At this moment, the first sampling took place: C and D plants were sampled from the first climate chamber, H, and D\*H plants were sampled from the second climate chamber. In order to perform a realistic experiment, corresponding with the dawn, heat exposure treatment began with an increasing temperature gradient from 20 to 40°C for 3 h, which was then maintained for 4 h. Lightweight expanded clay aggregate (LECA®), together with a refrigeration system, was used around the pots in order to mimic a fresher field soil temperature under heat stress. After that, environmental conditions inside the climate chambers were restored and all cuttings were well-watered (70% FC). The recovery of all groups was then evaluated at the second sampling point after 4 days under environmental and watering control conditions. In order to minimize the effects of environmental heterogeneity, the pots were periodically moved to the neighboring position during the whole experiment.

At each sampling point (first sampling point: stress; second sampling point: recovery), five plants per group (i.e., C, D, H, and D\*H) were used to evaluate plant water potential. Homogenous leaves from six individuals were used for *in vivo* measurements of leaf gas exchange parameters and subsequently used to determine plant relative water content (RWC) and electrolyte leakage. Also, homogeneous leaves from six individuals were immediately frozen in liquid nitrogen and kept at –80°C for further analysis (lipid peroxidation, redox couples ascorbate and glutathione, quantification of starch and pigments, hormonal alterations and metabolomics).

## Water Relations

Predawn water potential ( $\Psi_{pd}$ ) was measured using a Scholander-type pressure chamber (PMS Instrument Co., Corvallis, OR, United States). Four leaf discs (diameter = 11 mm) per individual were also collected to determine RWC, by using the equation:  $\text{RWC} = (\text{FW} - \text{DW}) / (\text{TW} - \text{DW}) \times 100$ , in which FW is the fresh weight, TW is the turgid weight after rehydration of the leaf discs for 24 h at 4°C in the dark, and DW is the dry

weight after oven-drying the leaf discs at 70°C until they reached a constant weight.

## Gas Exchange and Stomatal Conductance

Leaf gas exchange measurements were performed on fully expanded leaves using an infrared gas analyzer, LCpro-SD (ADC BioScientific Ltd., United Kingdom), equipped with the broad leaf chamber. Measurements were performed maintaining the following conditions inside the chamber: ambient temperature, CO<sub>2</sub> and H<sub>2</sub>O concentration, air flow  $200 \mu\text{mol s}^{-1}$  and light intensity  $400 \mu\text{mol m}^{-2} \text{s}^{-1}$ . Data were recorded when the measured parameters were stable (2–6 min). Net CO<sub>2</sub> assimilation rate (A), transpiration rate (E), stomatal conductance ( $g_s$ ), and internal CO<sub>2</sub> concentration ( $C_i$ ) were determined. Water use efficiency (WUE) was calculated based on leaf gas exchange, using the formulae  $\text{WUE} = A/E$ .

## Starch Quantification

Starch concentration was determined by using the anthrone method. Total soluble sugars were extracted from 50 mg of frozen leaves in 80% (v/v) ethanol for 1 h at 80°C. After centrifugation, the pellet was used to quantify starch, as described by Osaki et al. (1991). The pellet was resuspended with 30% (v/v) perchloric acid and incubated at 60°C for 1 h. The mixture was then centrifuged and anthrone was added to the supernatant. After heating the mixture at 100°C for 10 min, absorbance was read at 625 nm (Thermo Fisher Scientific Spectrophotometer, Genesys 10-uv S) and starch concentration was determined according to a D-glucose standard curve.

## Pigments Quantification

Concentration of chlorophyll a, b, and carotenoids was determined according to Sims and Gamon (2002). Pigments were extracted using cold acetone:50 mM Tris buffer pH 7.8 (80:20) (v/v). Following centrifugation, supernatant absorbance was read at 470, 537, 647, and 663 nm (Thermo Fisher Scientific Spectrophotometer, Genesys 10-uv S). Chlorophyll a, b, and carotenoids were then quantified by using the formulae presented by the author.

## Electrolyte Leakage and Lipid Peroxidation

To determine electrolyte leakage (EL), four leaf discs (diameter = 11 mm) were collected. Conductivity was measured (CONSORT C830, Consort bvba, Turnhout, Belgium) and EL was determined using the equation:  $\text{EL} = (C_i - W_c) / (C_f - W_c) \times 100$ , in which  $W_c$  represents water conductivity,  $C_i$  is the initial conductivity of water plus the leaf discs, and  $C_f$  is the final conductivity of water plus the leaf discs after 5 min at 121°C and 24 h at 4°C.

The extent of lipid peroxidation on leaves was estimated by measuring the amount of malondialdehyde (MDA), following an adaptation of the procedure described by Hodges et al. (1999). About 100 mg of leaves were ground in 2.5 mL of cold 0.1% (w/v) trichloroacetic acid (TCA) and centrifuged. A 250  $\mu\text{L}$  aliquot of

the supernatant was added to 1 mL of 20% (w/v) TCA containing 0.5% (w/v) TBA (positive control), and another 250  $\mu$ L was added to 1 mL of 20% (w/v) TCA (negative control). Both positive and negative controls per sample were heated at 95°C for 30 min. After stopping the reaction on ice, absorbance was read at 440, 532, and 600 nm (Thermo Fisher Scientific Spectrophotometer, Genesys 10-uv S, Waltham, MA, United States), and MDA content was determined by the formulae presented by the author.

## Non-protein Redox Couples Ascorbate and Glutathione

Ascorbate (reduced, AsA) and dehydroascorbate (DHA) concentrations, as well as oxidized (GSSG) and reduced (GSH) glutathione were determined according to the microplate method described by Queval and Noctor (2007).

## Hormone Quantification

Abscisic acid (ABA), jasmonic acid (JA), and salicylic acid (SA) were extracted and analyzed following the procedure described by Durgbanshi et al. (2005), with slight modifications. Freeze-dried tissue (50 mg) was mixed with 100 ng of ABA<sub>d6</sub>, 100 ng of SA<sub>d6</sub> and 100 ng of dihydrojasmonic acid and homogenized with 5 mL of distilled water. After cold centrifugation, supernatants were recovered and pH adjusted to 3 with 30% acetic acid. The acidified water extract was partitioned twice against 3 mL of diethyl ether. The organic upper layer was recovered and vacuum evaporated in a centrifuge concentrator (SpeedVac, Jouan, Saint Herblain, France). The dry residue was then resuspended in a 10% methanol solution by gentle sonication. The resulting solution was passed through 0.22  $\mu$ m regenerated cellulose membrane syringe filters (Albet S.A., Barcelona, Spain) and directly injected into a UPLC system (Acquity SDS, Waters Corp., Milford, MA, United States). Analytes were separated by reversed-phase (Nucleodur C18, 1.8  $\mu$ m 50  $\times$  2.0 mm, Macherey-Nagel, Barcelona, Spain) using a linear gradient of ultrapure water (A) and methanol (B) (both supplemented with 0.01% acetic acid) at a flow rate of 300  $\mu$ L min<sup>-1</sup>. The gradient used was: (0–2 min) 90:10 (A:B), (2–6 min) 10:90 (A:B) and (6–7 min) 90:10 (A:B). Hormones were quantified with a Quattro LC triple quadrupole mass spectrometer (Micromass, Manchester, United Kingdom) connected online to the output of the column through an orthogonal Z-spray electrospray ion source. The analytes were quantified after external calibration against the standards.

## Metabolomics Analysis

Metabolites were extracted, derivatized and analyzed by gas chromatography-mass spectrometry (GC-MS), as previously described by Foito et al. (2013). *Eucalyptus* leaves were lyophilized and 100 mg of dried, powdered material were weighed into glass tubes. Lyophilized material was extracted sequentially in methanol, water and chloroform for 30 min at 30°C each. Internal standards (aqueous ribitol and methanolic *n*-non-adeanoic acid) were added during the initial methanol extraction step. Finally, an additional aliquot of water was added and the polar and non-polar phases were separated, evaporated to dryness and derivatized independently. Metabolite

profiles of the polar and non-polar fractions were acquired following separation of compounds on a DB5-MSTM column (15 m  $\times$  0.25 mm  $\times$  0.25  $\mu$ m; J&W, Folsom, CA, United States) using a Thermo-Finnigan DSQ II GC-MS system (Thermo Finnigan, United Kingdom). The samples were analyzed as a single batch, in a randomized order, while quality control samples as well as blanks were incorporated at the beginning and the end of the sequence. Peak areas were calculated in relation to respective internal standard and normalized to respective extracted weight. Metabolites were identified based on their mass spectral characteristics and GC retention times, by comparison with retention times of reference compounds from an in-house reference library as previously described (Correia et al., 2016b).

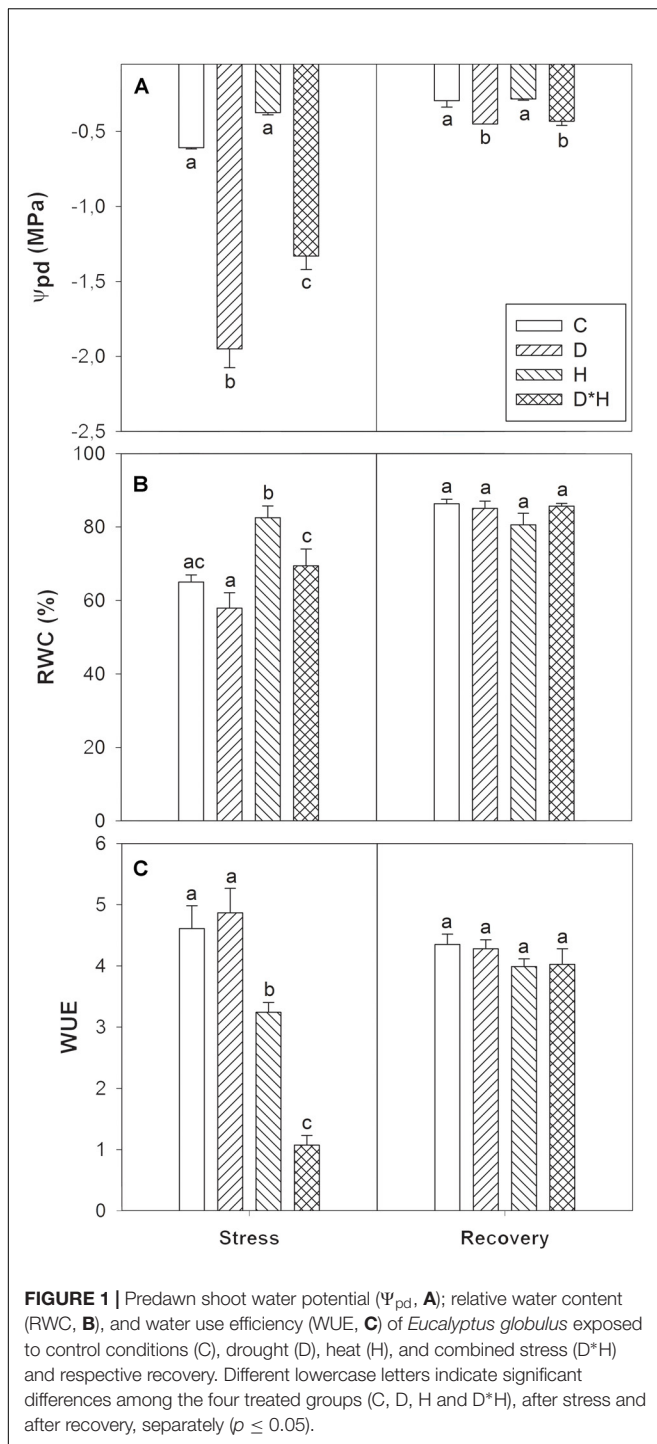
## Statistical Analysis

Data are presented as mean  $\pm$  SE (standard error) of three to six independent biological replicates. Statistical procedures were performed using SigmaPlot for Windows v. 11.0 (Systat Software Inc., San Jose, CA, United States), except metabolites that were analyzed using GenStat v16 (VSN International Ltd., Hemel Hempstead, United Kingdom). One-way analysis of variance (ANOVA) followed by the Fisher's LSD *post hoc* all pairwise multiple comparison tests were employed separately for each sampling point (i.e., stress and recovery) to estimate the significance of the results. Different lower cases indicate significant differences between treatments (C, D, H, and D\*H) at  $p \leq 0.05$ . In order to integrate the results, a complete dataset comprising all physiological, biochemical and metabolomic data was subjected to principal component analysis (PCA), sparse partial least squares (sPLS) and network analyses using the software R v3.1.2 core functions (R Core Team, 2014) plus the package mixOmics (Lê Cao et al., 2016). For building the sPLS model, the performance was first evaluated over 10 components, and two components (total Q2 > 0.1) were selected. Variables were then selected according to individual Q2, and those variables with a value lower than 0.1 were filtered out to prevent later overfitting of the model. The network was plotted employing the mixOmics network function, establishing a cut-off of 0.65 (which roughly correspond to plotting the variables with a Q2 > 0.35 for at least one of the two components).

## RESULTS

The effect of drought (D) and heat (H) stress applied alone and combined (D\*H) in *E. globulus* plants was analyzed by assessing physiological, biochemical, hormonal and metabolomic alterations after stress imposition and recovery. The plant water status was evaluated by  $\Psi_{pd}$ , RWC and WUE. Drought and combined stress induced a significant reduction in  $\Psi_{pd}$ , with the extent of reduction being higher in the drought stress alone than in the combined stress (Figure 1A). After recovery, although increased,  $\Psi_{pd}$  of drought and combined stress was still lower than the control (Figure 1A). RWC was only slightly decreased after the drought treatment (not significant), significantly increased in heat, and was unaffected in the combined stress (Figure 1B). Finally, WUE revealed





a reduction in heat and combined stress, with the combined stress exhibiting the most severe reduction in WUE (**Figure 1C**). Both RWC and WUE fully recovered after returning to control conditions (**Figures 1B,C**).

Gas exchange varied in response to the imposed stresses (**Figure 2**). Net photosynthetic rate (A) was reduced following all stress treatments (**Figure 2A**) with the combined stress leading to the greatest reduction, followed by drought and heat

stress. Transpiration rate (E) and stomatal conductance ( $g_s$ ) were similarly affected with only drought and combined stress resulting in a decrease (**Figures 2B,C**). On the other hand, internal  $CO_2$  concentration ( $C_i$ ) significantly increased in drought, and decreased in heat, remaining unchanged in the combined stress (**Figure 2D**). Most of these responses only slightly leveled off following recovery (**Figures 2A–C**).

Leaf starch content increased in drought, decreased in heat and slightly although not significantly decreased in the combined stress. After recovery leaf starch was similar regardless of prior treatment (**Figure 3**).

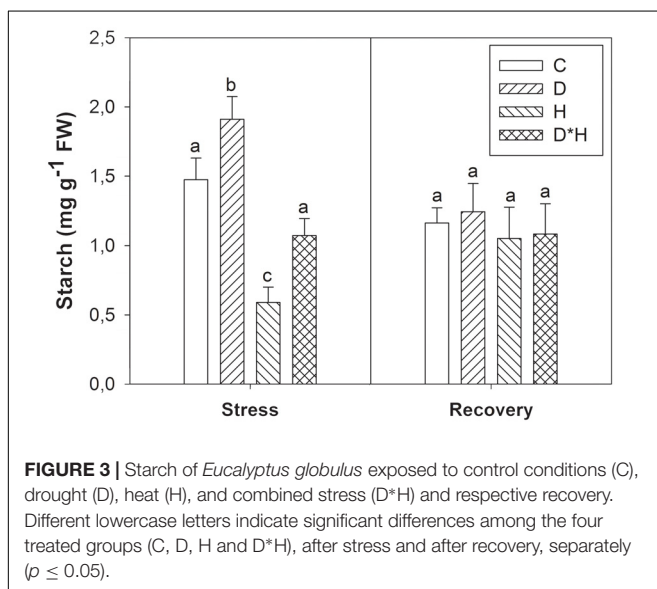
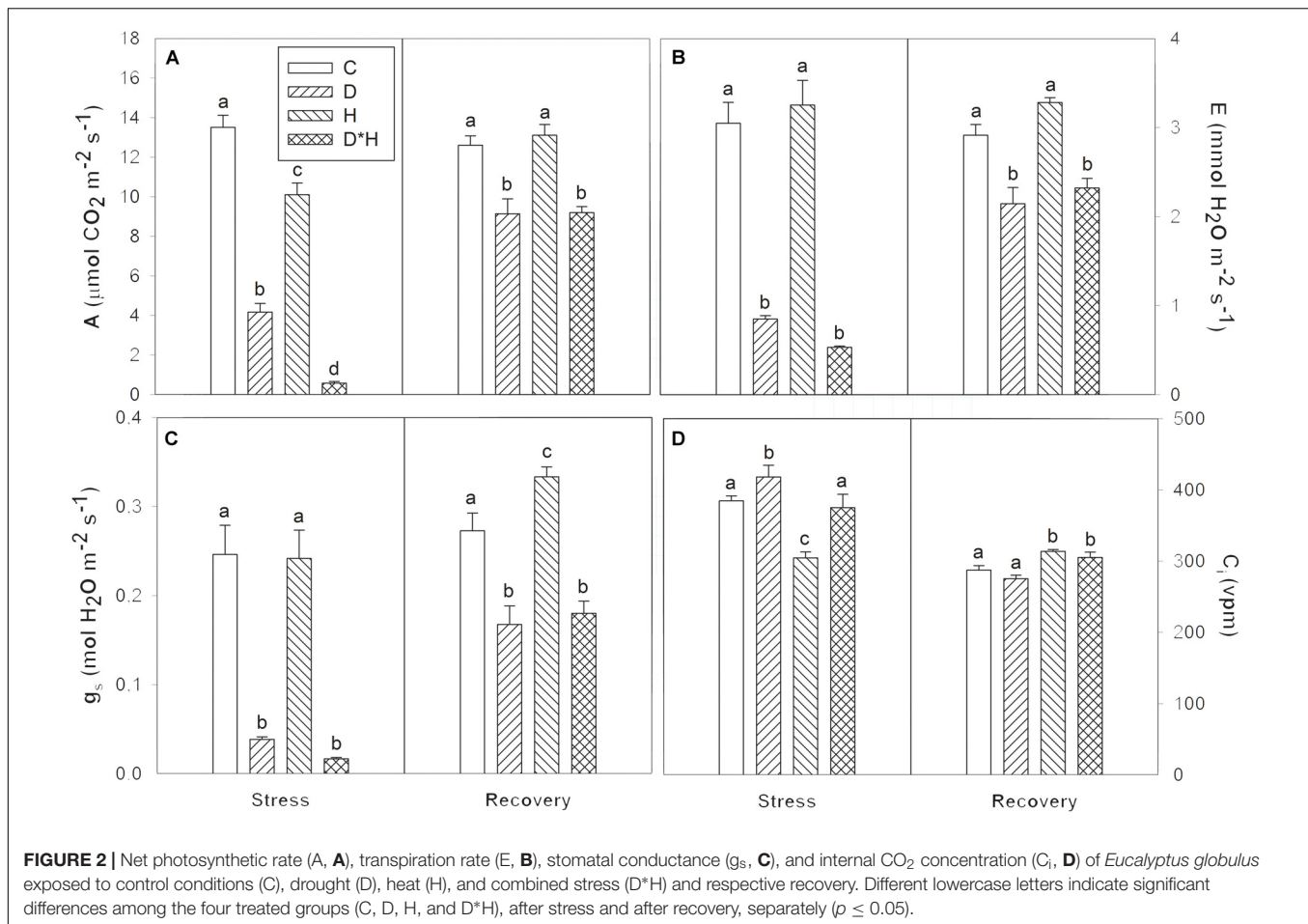
Chlorophyll a, b, and carotenoids were differentially modulated by the imposed stresses (**Figure 4**). Chlorophyll a decreased in drought and heat (**Figure 4A**) while chlorophyll b was only reduced after the heat treatment (**Figure 4B**). After recovery, chlorophyll a was higher in previously drought stressed plants than in control plants (**Figure 4A**). The carotenoid abundance profile matched chlorophyll b exhibiting a major reduction only in the heat stress and an increase in drought stressed plants after recovery (**Figure 4C**).

Membrane integrity was assessed by leaf EL and MDA accumulation. Electrolyte leakage revealed a significant increase during drought stress (**Figure 5A**) that was accompanied by a trend towards higher MDA at the same point (**Figure 5B**). After recovery, all plants exhibited equivalent EL and MDA content (**Figures 5A,B**).

The total AsA pool was not affected by the imposed stresses; on the contrary the total glutathione pool was increased in the drought and combined stress (**Figure 6**). This induction was not accompanied by an increase in the oxidized pool (**Figure 6**). After recovery, an increase in the oxidation status of AsA pool was observed without major alterations in the total AsA content. Glutathione content of leaves subjected to drought or combined stress return to control levels following recovery (**Figure 6**).

The imposed stresses significantly affected the leaf hormonal dynamics and major differences were found regarding ABA, SA, and JA, as shown in **Figure 7**. On one hand, ABA significantly accumulated in drought and combined stress (**Figure 7A**). On the other hand, SA levels decreased exclusively after heat (**Figure 7B**) and JA content decreased under all stress conditions in a descending order: drought, heat, and combined stress (**Figure 7C**). No hormonal differences were detected after recovery from stress.

The foliar metabolite profile of *E. globulus* subjected to drought, heat, combined stress and recovery were compiled using GC–MS. This analysis yielded the detection of 106 metabolites (Supplementary Table S1), distinguishing between 64 polar and 42 non-polar metabolites. Only a small part of the detected metabolites could not be identified after data processing (7 polar and 5 non-polar metabolites). From the identified metabolites, 48 showed significant changes due to the applied stress and/or recovery, including 12 carbohydrates (**Table 1**), 5 organic acids and 17 amino acids (**Table 2**), 2 phenolic acids, 6 fatty acids/alcohols, 1 phytosterol, and 5 unknown metabolites (**Table 3**).

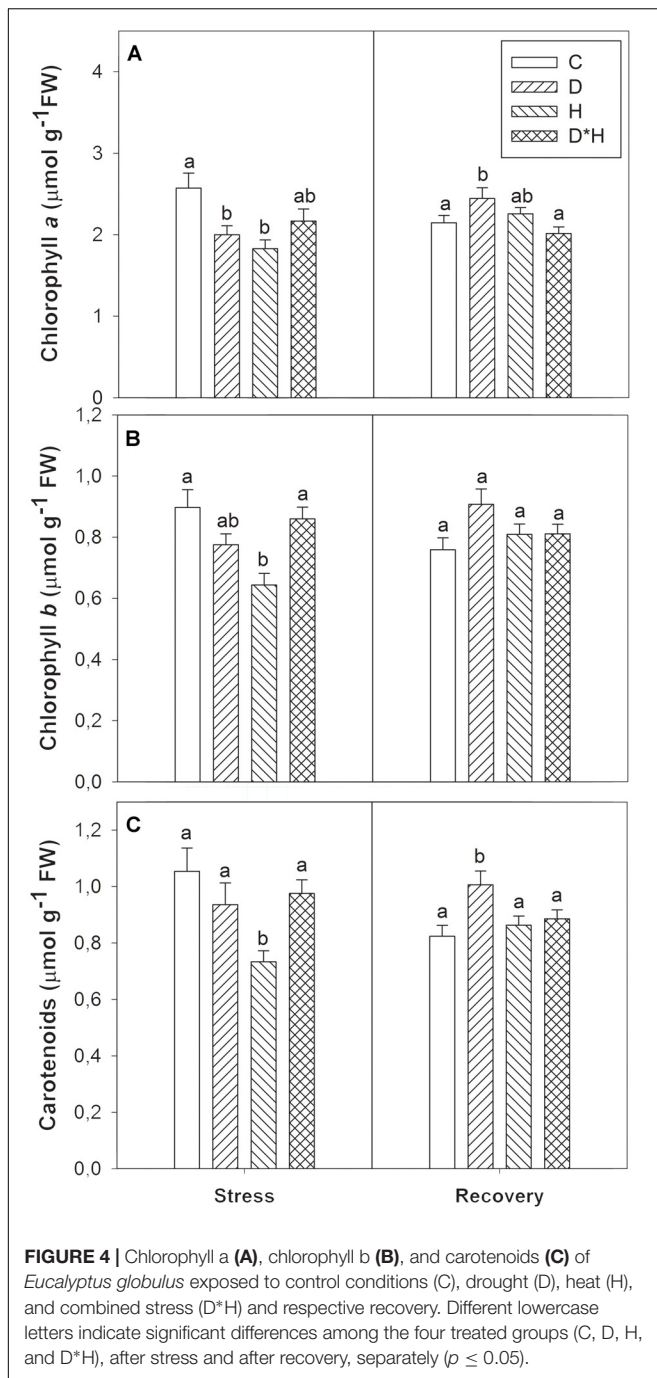


Regarding carbohydrates (**Table 1**), mannose, galactose and two separate peaks assigned to glucose increased exclusively after drought. Mannitol, sorbitol and inositol contents increased

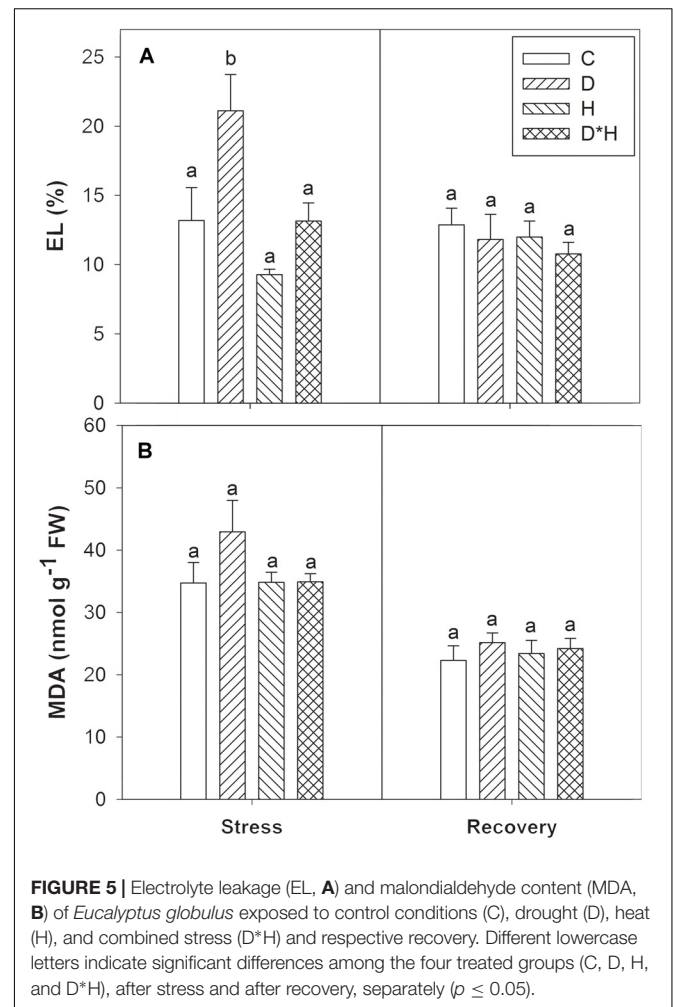
under all stress conditions, and maltose increased in drought and combined stress. On the contrary, abundances of fructose-6-phosphate and glucose-6-phosphate were negatively affected by drought, heat, and combined stress. Dihydroxydihydrofuranone also exhibited decreases under all stress conditions, but with lower magnitude. After recovery, both fructose-6-phosphate and glucose-6-phosphate of heat and combined stressed plants reversed to control levels, but drought stressed plants still kept significantly lower content (**Table 1**). All other carbohydrates reversed the alterations caused by stress after recovery, except inositol that maintained higher levels in plants previously exposed to drought and combined stress (**Table 1**).

Five organic acids – succinate, malate, citrate, quinate, and glycerate – were modulated by stress (**Table 2**). Succinate, quinate, and glycerate abundances were reduced under both drought and combined stress. From these, glycerate was also decreased under heat. Citrate was elevated in response to drought, heat, and combined stress (**Table 2**). Malate abundance was enhanced by drought and reduced after heat, staying unchanged in the combined stress. Following recovery, none of the organic acids showed significant changes relative to control plants (**Table 2**).

Amino acids constitute the largest group of compounds showing significant differences under stress, mainly combined

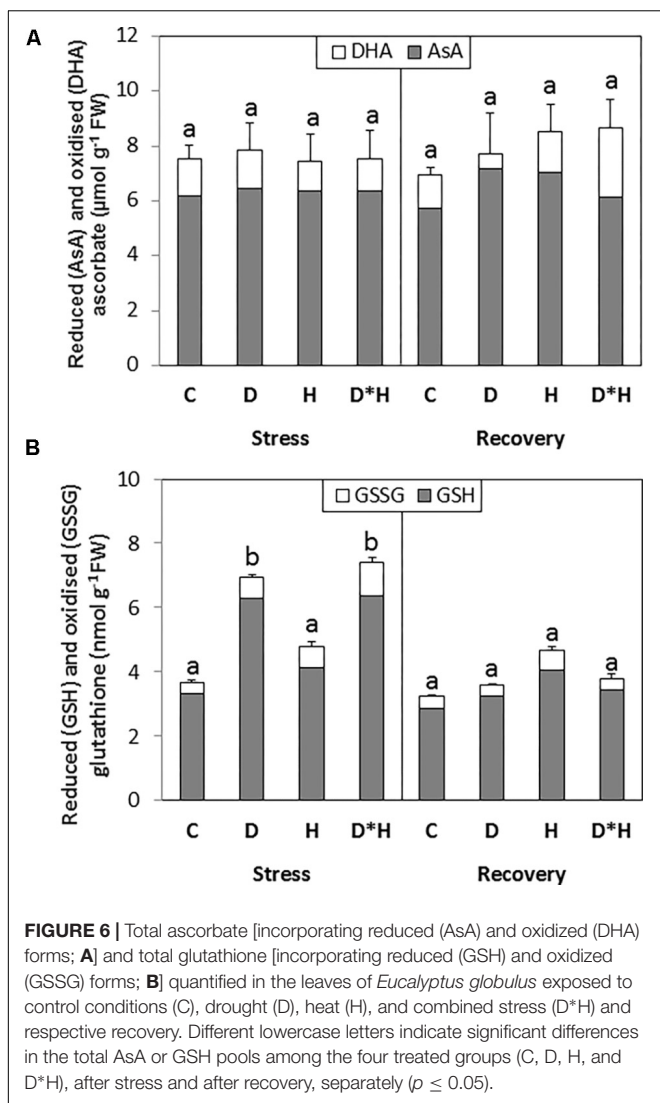


stress (Table 2). Aspartate, glutamate, leucine, isoleucine, and proline abundances were significantly increased in drought and heat, and to a greater extent, in the combined stress. Threonine, lysine, histidine, tryptophan, methionine, and GABA were only positively regulated under drought and combined stress, and valine showed an over accumulation only in the combined stress. After recovery, only oxoproline (generated from glutamine during the derivatization procedure), tryptophan and methionine revealed significant alterations (Table 2). Oxoproline decreased in previously droughted and combined stressed plants,



tryptophan could only be detected in previously drought stressed plants, and methionine decreased in the combined stress, although showing a slight decrease in previously heat stressed plants. Drought, heat, and combined stress positively induced urea levels, which were restored after recovery (Table 2).

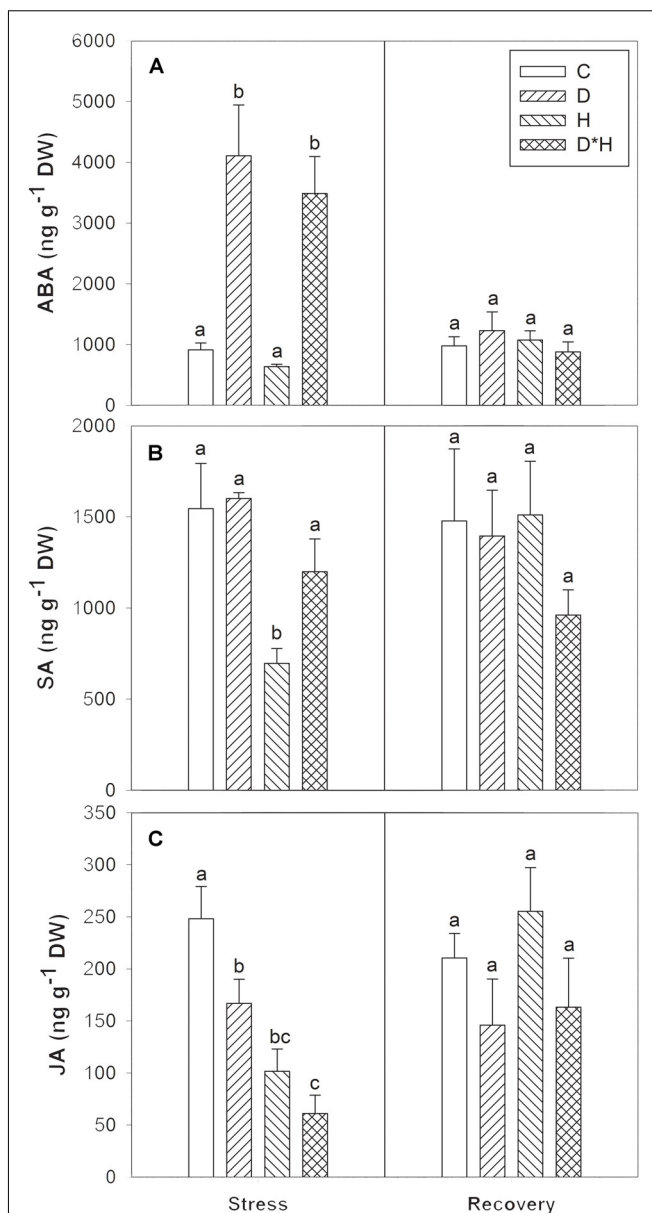
Regarding phenolic acids (Table 3) shikimate decreased under drought and combined stress. Conversely, cinnamate abundance increased under combined stress, although a slight increase was also observed in drought. Neither of these phenolic acids showed significant differences after recovery. Alterations in fatty acids and fatty alcohols were mainly detected in recovery (Table 3), with the exception of C14:0, which increased under every stress, and C18:2, which slightly reduced under drought and combined stress. After recovery, C14:0 abundance in previously stressed plants largely reversed to control levels and only plants previously subject to combined stress still presented a significantly higher content. Recovery from drought induced C23:0 and C29:0 accumulation; recovery from heat reduced C21:0 and C26 alcohol; and recovery from combined stress resulted in enhanced C29:0 and reduced C26:0 alcohol. Avenasterol was the only identified phytosterol showing significant changes (Table 3); however, the difference observed is due to a lower level of



avenasterol on drought in relation to heat and combined stress, representing only a slight decrease when compared with control.

From the unidentified metabolites (Table 3), pU2020 increased under drought and pU1585 raised under combined stress, pU1598 and pU2367 increased under both conditions. Moreover, npU1680 revealed an increase after recovery from combined stress.

The supervised and unsupervised integrated analysis provided us a comprehensive overview of the plant stress responses, identifying the most relevant interactions. Initial comparison based on principal component analysis (PCA) exhibited a clear separation between control and differentially stressed samples in *E. globulus* (Figure 8). Control sample scores were grouped together at the bottom left quadrant. Sample scores of plants subject to heat were located more to the right and downwards compared to control, whereas drought sample scores were also placed on the right quadrant but upwards. The combined sample scores were all found together on the bottom right quadrant, farther from the control scores compared to the other stress



conditions. The separation between control and heat was mainly related to increased levels of phenylalanine, RWC and some non-polar metabolites, whereas drought separation is mainly driven by accumulation of several sugars (fructose, mannose, glucose), starch, ABA, MDA, and glutathione (Supplementary Table S2). On the other hand, the most separated condition, combined, is mainly driven by increased phenylalanine, RWC and some non-polar metabolites, as seen in heat samples, accumulation of starch, ABA, MDA, and glutathione, as seen in drought samples,



TABLE 1 | Metabolomic analysis, relative abundance of carbohydrates.

Carbohydrates	Relative abundance (C – D – H – D*H)	
	Stress	Recovery
Mannose	a b a a	-----
	■ ■ _ _	
Galactose	a b a a	-----
	_ ■ _ _	
Glucose 1	a b a a	-----
	_ ■ _ _	
Glucose 2	a b a a	-----
	_ ■ _ _	
Maltose	a b a b	-----
	_ ■ _ ■	
Mannitol	a b b b	-----
	_ ■ ■ ■	
Sorbitol	a b b b	-----
	_ ■ ■ ■	
Inositol	a b c b	a b a b b
	_ ■ _ ■	_ ■ ■ ■
Fructose-6-phosphate	a b c b	a b a a b
	■ _ _ _	■ _ _ ■
Glucose-6-phosphate	a b a b b	a b a a b
	■ _ _ _	■ _ ■ ■
Dihydroxydihydrofuranone	a b b b	-----
	■ ■ ■ ■	
α-Glycerophosphate	ab a a b	-----
	_ _ _ ■	

Abundance data is presented on a scale relative to the lowest value among treatments. Different lowercase letters indicate significant differences between the control (C), drought (D), heat (H), and combined stress (D\*H) plant groups, under stress and after recovery; absent relative abundance indicates that no differences were found ( $p \leq 0.05$ ).

together with higher levels of several amino acids (leucine, isoleucine, histidine, tryptophan, asparagine; Supplementary Table S2). After recovery, these differences were mostly reversed and the sample scores of the stressed conditions were found near control sample scores (Figure 8).

The constructed network based on sparse partial least squares (sPLS) allowed the determination of the specific components behind the observed phenotypical changes considering the complete metabolomic, biochemical and physiological changes of the experiment (Figure 9). This network highlighted photosynthesis (A), MDA, glutathione (GSH.GSSG) and ascorbate (AsA.DHA) as biochemical central points, which are positively and negatively correlated with the studied hormones (ABA and JA) through several metabolomic alterations (Figure 9). Specifically, JA and AsA.DHA are negatively correlated with glutamic acid, isoleucine, lysine, aspartic acid, proline, and tryptophan (Figure 9). Most of these are positively correlated with glutathione and MDA and also mediating a positive correlation between these and other amino acids (GABA, histidine, leucine), and ABA. Finally, another key interaction places A with a positive relation with putrescine, glucose-6-phosphate, fructose-6-phosphate, and quinic acid

TABLE 2 | Metabolomic analysis, relative abundance of organic acids and amino acids.

Organic acids	Relative abundance (C – D – H – D*H)	
	Stress	Recovery
Succinate	a b a a	-----
	■ ■ ■ ■	
Malate	a b c a	-----
	_ ■ _ _	
Citrate	a b b b	-----
	_ ■ ■ ■	
Quinate	a b c a b	-----
	■ _ ■ _	
Glycerate	a b c c	-----
	■ _ _ _	
<b>Amino acids</b>		
Aspartate	a b b c	-----
	_ ■ ■ ■	
Glutamate	a b b c	-----
	_ ■ ■ ■	
Asparagine 1	a b c a b c	-----
	_ _ _ ■	
Asparagine 2	a b c a b a	-----
	_ _ _ ■	
Oxoproline		a b c a b a
	-----	■ _ ■ ■
Valine	a a a b	-----
	_ _ _ ■	
Leucine	a a b a c	-----
	_ _ _ ■	
Isoleucine	a b b c	-----
	_ ■ ■ ■	
Proline	a b b c	-----
	_ ■ ■ ■	
Threonine	a b a b	-----
	_ _ _ ■	
Lysine	a b a c	-----
	_ _ _ ■	
Histidine	a b a b	-----
	_ _ ■	
Phenylalanine	a a b b	-----
	_ _ ■ ■	
Tryptophan	a b b b	a b a a
	_ _ ■	■
Methionine	a b a b	a a a b b
	_ _ _ ■	■ ■ _ _
GABA	a b a b	-----
	_ ■ _ ■	
Urea	a b b b	-----
	_ _ _ ■	

Abundance data is presented on a scale relative to the lowest value among treatments. Different lowercase letters indicate significant differences between the control (C), drought (D), heat (H), and combined stress (D\*H) plant groups, under stress and after recovery; absent relative abundance indicates that no differences were found ( $p \leq 0.05$ ).

**TABLE 3 |** Metabolomic analysis, relative abundance of phenolic acids, fatty acids/alcohols, phytosterols, and unknown metabolites.

Phenolic acids	Relative abundance (C – D – H – D*H)	
	Stress	Recovery
Shikimate	a b a b	-----
	■ ■ ■	
Cinnamate	a ab a b	-----
	■ ■ ■ ■	
<b>Fatty acids, fatty alcohols</b>		
C14:0	a b b b	a b b b
	■ ■ ■ ■	■ ■ ■ ■
C18:2	ab a b a	-----
	■ ■ ■ ■	
C21:0	-----	ab a c b c
		■ ■ ■ ■
C23:0	-----	a b a a
		■ ■ ■ ■
C29:0	-----	a b a b
		■ ■ ■ ■
C26 alcohol	-----	a ab b c c
		■ ■ ■ ■
<b>Phytosterols</b>		
Avenasterol	-----	ab a b b
		■ ■ ■ ■
<b>Unknowns</b>		
pU1585	a a a b	-----
	■ ■ ■ ■	
pU1598	a b a b	-----
	■ ■ ■ ■	
pU2020	a b a a	-----
	■ ■ ■ ■	
pU2367	a b a b	-----
	■ ■ ■ ■	
npU1680	-----	a b a a
		■ ■ ■ ■

Abundance data is presented on a scale relative to the lowest value among treatments. Different lowercase letters indicate significant differences between the control (C), drought (D), heat (H), and combined stress (D\*H) plant groups, under stress and after recovery; absent relative abundance indicates that no differences were found ( $p \leq 0.05$ ).

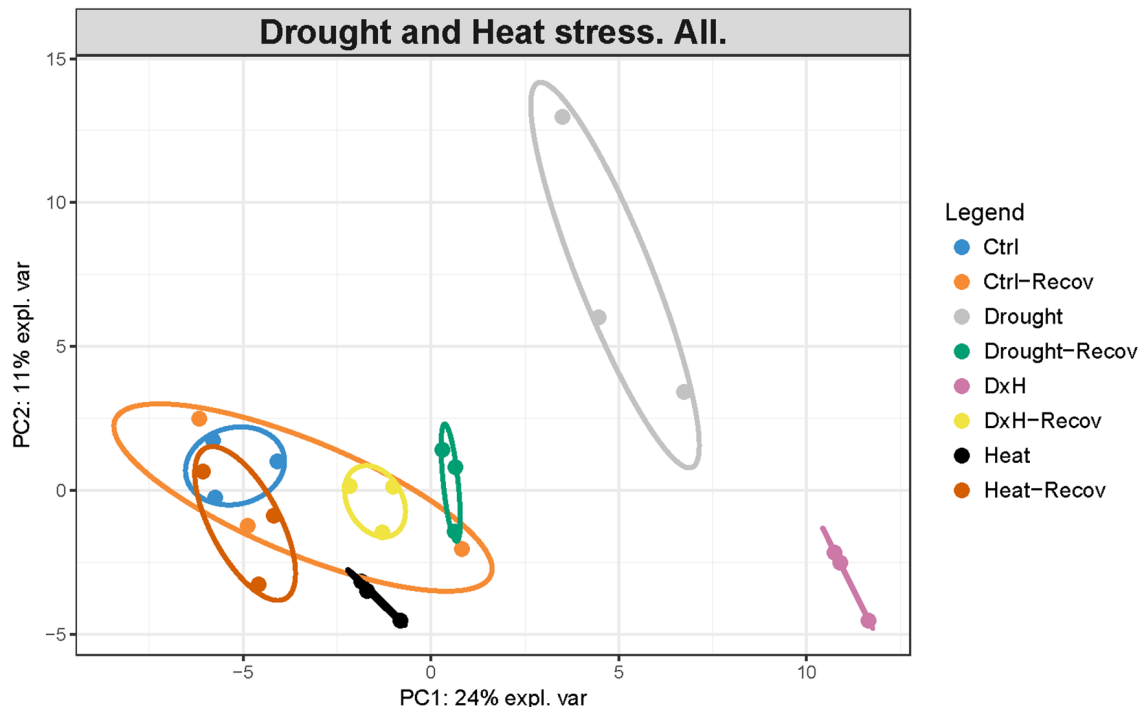
(Figure 9). This last one, quinic acid, also mediates a positive interaction between A, on one side, and  $\Psi_{pd}$ , E and  $g_s$ , on the other. All of these reveal negative interactions with most of the intervening amino acids.

DISCUSSION

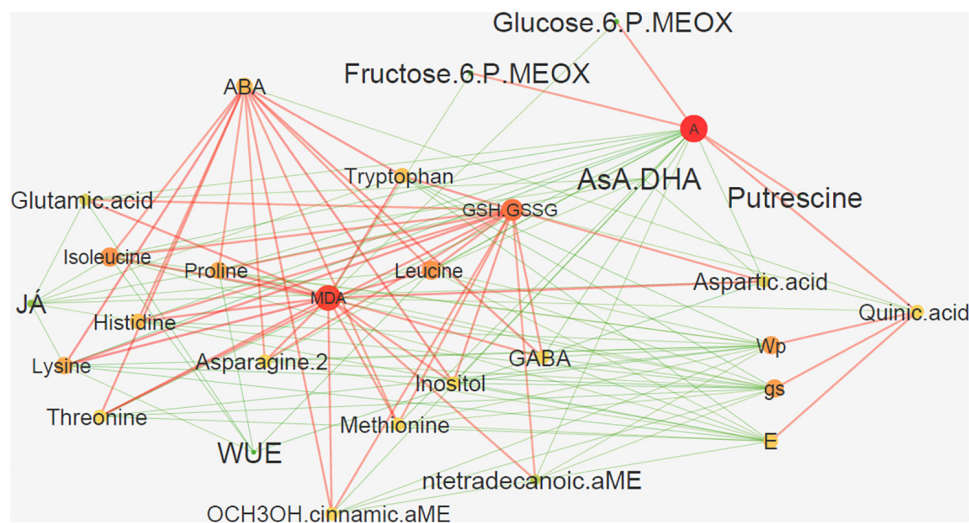
Considerable research advances have been accomplished focusing on plant responses to single stress factors under controlled environments (Mittler and Blumwald, 2010). However, plants growing in the field encounter a number of different co-occurring abiotic stresses that most probably cannot be extrapolated by the sum of the different stresses applied individually, altering plant metabolism in a novel manner (Rizhsky et al., 2002;

Zandalinas et al., 2016). Bearing this in mind, we aimed to determine convergent and divergent responses of the individual stresses in relation to their combination, evaluating the impact of drought and heat stress (alone and in combination) and respective recovery using a drought-tolerant *E. globulus* clone. Regarding drought stress alone, the main responses included reduced  $\Psi_{pd}$ , gas exchange, JA, fructose-6-phosphate, glucose-6-phosphate,  $\alpha$ -glycerophosphate, and shikimate, and increases in MDA, glutathione pool, ABA, amino acids, starch, and non-structural carbohydrates. Most of these results are in agreement with other reports that analyzed the isolated effect of drought on *E. globulus* (Warren et al., 2011; Correia et al., 2014b, 2016a,b), and indicate that water deficit negatively affects plant water relations and photosynthesis, causing a moderate oxidative stress, and inducing enhanced osmoprotection and other defence-related pathways. On the other hand, heat stress alone triggered an increase in RWC,  $\Psi_{pd}$ , mannitol, sorbitol, inositol and several amino acids that were accompanied by a reduction in the photosynthetic rate and pigments, WUE, starch, fructose-6-phosphate, glucose-6-phosphate,  $\alpha$ -glycerophosphate, SA, and JA. The reduction in the photosynthetic rate and pigments in parallel with unaffected transpiration rate and stomatal conductance confirms the particular sensitivity of photosynthesis to heat stress (Sharkey, 2005), even in a short heat shock (4 h at 40°C). It also indicates that the main limitations are non-stomatal and mostly related to heat-induced alterations in enzyme activity (Larkindale et al., 2005). A decrease in photosynthetic pigments, fructose-6-phosphate, glucose-6-phosphate, and starch was also documented in potato leaves growing at a moderately elevated 30°C (Hancock et al., 2014). No major oxidative impairment was detected and this can be explained by the shifts in the polyols mannitol, sorbitol, inositol and several amino acids, such as proline, possibly indicating that these compatible solutes were effective hydroxyl radical scavengers (Smirnoff and Cumbe, 1989; Wang et al., 2003). In addition to their role as radical scavengers, the accumulation of the polyols under heat stress is most likely responsible for the observed increase in RWC and  $\Psi_{pd}$ , reinforcing their primary role as osmoprotectants (Bokszczanin et al., 2013).

The heat-induced reduction in SA and JA is an unexpected result since both hormones are reported to play an important role as signal molecules in abiotic stress tolerance (Horváth et al., 2007; Xu et al., 2016) and SA has been reported to protect plants from heat stress (Wang et al., 2014). However, the downregulation of JA has already been described in *E. globulus* under water deficit (Correia et al., 2014a). Our results further confirm the downregulation of JA under drought stress, highlighting a similar response triggered by heat stress regarding not only JA but also SA. The way these abiotic stresses influence these two phytohormones in *E. globulus* is yet to be discovered. However, SA and JA, together with ethylene, are known to play major roles in regulating plant defense. SA is usually associated with the activation of defense against biotrophic and hemi-biotrophic pathogens, and the establishment of systemic acquired resistance (SAR). JA and ethylene are generally involved in defense against necrotrophic pathogens and herbivorous insects



**FIGURE 8 |** Principal Component Analysis (PCA) of a complete dataset of physiological, biochemical and metabolomic alterations occurring in *Eucalyptus globulus* after several stress conditions (drought, heat, and combined) and recovery. First two components are plotted in the graph. The proportion of variance explained by each component is indicated on axis labels.



**FIGURE 9 |** sPLS-based interaction network of the major alterations occurring in *Eucalyptus globulus* after several stress conditions (drought, heat, and combined) and recovery. Network is presented using the force-directed layout, which is based on the “force-directed” paradigm. Color and size of nodes illustrate radiality on a small to large and green to red scale (radiality of a node is calculated by computing the shortest path between the node and all other nodes in the graph); color and size of edges reflect weight with red and green edges illustrating positive and negative correlations, respectively.

[reviewed by Bari and Jones (2009)]. Hence, this result has significance in terms of the impact of abiotic stress on biotic interactions, suggesting that these abiotic stresses can negatively influence defense against other biotic threats.

A divergent response between isolated drought and heat stress is related to changes in the TCA cycle intermediates. In heat, citrate increase went along with reduced malate, whereas drought-induced increases of citrate and malate were

accompanied by reduced succinate. Together with the different amino acids that accumulate in each stress this result highlights two different metabolic regulations. In heat, the TCA cycle flux appears to be changed to two weakly connected branches, with malate functioning as a mitochondrial respiratory substrate to produce citrate, which is then converted to glutamate and proline. Similar cases of the non-cyclic flux mode of TCA cycle has been reviewed elsewhere (Sweetlove et al., 2010). However, the prevailing pathway activated under this condition appears to be the shikimic acid pathway, revealed by the over accumulation of shikimate and phenylalanine. Conversely, the shikimic acid pathway is downregulated under drought conditions. In this stress scenario, an induction in the first steps of the TCA cycle likely supplies higher demands for citrate that is metabolized to amino acids of the glutamate family; and succinate is converted to malate, which in turn is redirected to produce amino acids of the oxaloacetate/aspartate family.

Still on this subject, comparing the isolated stresses with the combined one reveals novel responses. In the combined conditions of drought and heat stress, the highest accumulation of citrate was accompanied by reduced succinate without major alterations in malate. The higher content of  $\alpha$ -glycerophosphate together with the major accumulation in amino acids of the glutamate family, the oxaloacetate/aspartate family and leucine/valine indicates that glycolysis is enhanced in this combined condition, sustaining the higher demand for amino acids. Still, we should also note the possibility for amino acid mobilization resulting from protein breakdown as protein turnover has been described as an important regulatory mechanism that allows plant cells to respond to drought and recovery (Lyon et al., 2016). The absence of significant changes in the fatty acids/alcohols and phytosterols detected at this new stress state does not support the premise of a regulation by changes of membrane lipids as we could assume (Falcone et al., 2004).

A novel response triggered only by the combined effect of drought and heat was the induction of cinnamate. Cinnamate originates all phenylpropanoids through the action of phenylalanine ammonia-lyase (PAL) on phenylalanine (Dixon and Paiva, 1995). We are yet uncertain of which phenylpropanoids are generated under this condition since a number of different phenylpropanoids can be involved (Dixon and Paiva, 1995).

Drought-stressed plants subject to a heat shock revealed a decrease in gas exchange (sharp), WUE,  $\Psi_{pd}$  and JA, no alterations in EL, MDA, starch and pigments and increased glutathione pool in relation to control. Comparing with drought stress alone, this reveals that subjecting drought stressed plants to an additional heat stress alleviated  $\Psi_{pd}$  and MDA, maintaining an increased glutathione pool and reducing starch content and non-structural carbohydrates. Interestingly, and in contrast to the expected negative effect of the stress combination on plant growth reported for other species (Silva et al., 2010), these results highlight that the combination of drought and heat provides significant protection from more detrimental effects of drought-stressed eucalypts. A similar conclusion has been

described for tomato plants under the combined effect of salinity and heat (Rivero et al., 2014).

Regarding recovery, most of the parameters affected by each stress condition reversed after re-establishment of the control growing condition. This is a common reported response (Correia et al., 2014a,b, 2016b; Escandón et al., 2016). Gas exchange and some carbohydrates reversed at a slower pace after drought and combined stress, which reveals the sensitivity of the photosynthetic apparatus (Chaves et al., 2009) and points out the most restrictive effect of these two stress conditions. On the other hand, the different modulation of several fatty acids/alcohols and phytosterols after recovery from drought and combined stress uncovers a putative regulation that allows restoration after stress through changes of membrane lipids (Falcone et al., 2004).

In accordance with the idea that relatively few studies have attempted to correlate metabolite content with physiological data, and the advantages of those (Tohge et al., 2015), we decided to introduce an integrative approach to analyze our dataset. The PCA and network results summarize the overall knowledge acquired in our study, aligning with some regulatory networks already described for their involvement in tolerance and recovery to drought (Brossa et al., 2015; Lyon et al., 2016), as well as other stresses (Das and Roychoudhury, 2014; Xu et al., 2015).

At present, information on the combined effect of heat and drought stress in *Eucalyptus* is rather limited although much needed from the application point of view (e.g., finding suitable markers for selecting the most tolerant genotypes to field establishment). In this work, we have reported different physiological, biochemical and metabolic adjustments that enable *E. globulus* to thrive under conditions of drought and heat applied alone or in combination. Although a few mechanisms were convergent to all stress conditions, the response magnitude was very dependent on the specific stress, and most of the metabolic pathways responded uniquely to each specific stress. Rather than presenting an additive outcome, the combination of heat stress ameliorated part of the negative effect of drought. The information collected here confirms that the biological processes switched on by an environmental factor are very specific to that exact condition and are likely to differ from those activated by a slightly different environmental condition (Mittler, 2006; Atkinson and Urwin, 2012). The need for studies that focus on the actual field stress conditions is thus evident and imperative for selecting plants with enhanced tolerance to naturally occurring environmental conditions.

## AUTHOR CONTRIBUTIONS

GP designed and supervised the experimental procedure. RDH designed and supervised the biochemical and metabolomic characterization. AG-C designed and supervised the hormonal quantification. BC, JA, and GP performed the experiment, the physiological, and the biochemical characterization. BC and RDH performed the metabolomic profiling and analysis. LV designed and performed



the PCA and network analysis. BC, RDH, and GP wrote the manuscript. All authors discussed the data and reviewed the manuscript.

## FUNDING

This research was supported by FEDER within the PT2020 Partnership Agreement and Compete 2020 (Programa Operacional Fatores de Competitividade) and by National Funds through the Portuguese Foundation for Science and Technology (FCT), which financed CESAM (UID/AMB/50017 - POCI-01-0145-FEDER-007638) and the project PTDC/AGR-CFL/112996/2009. FCT also supported the fellowships of BC

(SFRH/BD/86448/2012), JA (SFRH/BD/120967/2016), and GP (SFRH/BPD/101669/2014). LV was supported by the Ramón y Cajal Program (RYC-2015-17871) (Spanish Ministry of Economy and Competitiveness). The James Hutton Institute receives support from the Rural and Environmental Science and Analytical Services Division of the Scottish Government.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.00819/full#supplementary-material>

## REFERENCES

- Allen, C. D., Breshears, D. D., and McDowell, N. G. (2015). On underestimation of global vulnerability to tree mortality and forest die-off from hotter drought in the Anthropocene. *Ecosphere* 6:129. doi: 10.1890/ES15-00203.1
- Anderegg, W. R., Berry, J. A., Smith, D. D., Sperry, J. S., Anderegg, L. D., and Field, C. B. (2012). The roles of hydraulic and carbon stress in a widespread climate-induced forest die-off. *Proc. Natl. Acad. Sci. U.S.A.* 109, 233–237. doi: 10.1073/pnas.1107891109
- Arend, M., Brem, A., Kuster, T., and Günthardt-Goerg, M. (2013). Seasonal photosynthetic responses of European oaks to drought and elevated daytime temperature. *Plant Biol.* 15, 169–176. doi: 10.1111/j.1438-8677.2012.00625.x
- Atkinson, N. J., and Urwin, P. E. (2012). The interaction of plant biotic and abiotic stresses: from genes to the field. *J. Exp. Bot.* 63, 3523–3543. doi: 10.1093/jxb/ers100
- Bansal, S., Hallsby, G., and Nilsson, M.-C. (2013). Synergistic, additive and antagonistic impacts of drought and herbivory on *Pinus sylvestris*: leaf, tissue and whole-plant. *Tree Physiol.* 33, 451–463. doi: 10.1093/treephys/tp1019
- Bari, R., and Jones, J. D. G. (2009). Role of plant hormones in plant defence responses. *Plant Mol. Biol.* 69, 473–488. doi: 10.1007/s11103-008-9435-0
- Bokszczanin, K. L., Fragkostefanakis, S., Bostan, H., Bovy, A., Chaturvedi, P., Chiusano, M. L., et al. (2013). Perspectives on deciphering mechanisms underlying plant heat stress response and thermotolerance. *Front. Plant Sci.* 4:315. doi: 10.3389/fpls.2013.00315
- Booth, T. H. (2013). Eucalypt plantations and climate change. *For. Ecol. Manage.* 301, 28–34. doi: 10.1016/j.foreco.2012.04.004
- Brodribb, T. J., and Cochard, H. (2009). Hydraulic failure defines the recovery and point of death in water-stressed conifers. *Plant Physiol.* 149, 575–584. doi: 10.1104/pp.108.129783
- Brossa, R., Pintó-Marijuan, M., Francisco, R., López-Carbonell, M., Chaves, M. M., and Alegre, L. (2015). Redox proteomics and physiological responses in *Cistus albidus* shrubs subjected to long-term summer drought followed by recovery. *Planta* 241, 803–822. doi: 10.1007/s00425-014-2221-0
- Chaves, M. M., Flexas, J., and Pinheiro, C. (2009). Photosynthesis under drought and salt stress: regulation mechanisms from whole plant to cell. *Ann. Bot.* 103, 551–560. doi: 10.1093/aob/mcn125
- Chaves, M. M., Pereira, J. S., Maroco, J., Rodrigues, M., Ricardo, C. P. P., Osorio, M., et al. (2002). How plants cope with water stress in the field? Photosynthesis and growth. *Ann. Bot.* 89, 907–916. doi: 10.1093/aob/mcf105
- Correia, B., Hancock, R. H., Valledor, L., and Pinto, G. (2018). Gene expression analysis in *Eucalyptus globulus* exposed to drought stress in a controlled and a field environment indicates different strategies for short- and longer-term acclimation. *Tree Physiol.* doi: 10.1093/treephys/tpy067
- Correia, B., Pintó-Marijuan, M., Castro, B. B., Brossa, R., López-Carbonell, M., and Pinto, G. (2014a). Hormonal dynamics during recovery from drought in two *Eucalyptus globulus* genotypes: from root to leaf. *Plant Physiol. Biochem.* 82, 151–160. doi: 10.1016/j.plaphy.2014.05.016
- Correia, B., Pintó-Marijuan, M., Neves, L., Brossa, R., Dias, M. C., Costa, A., et al. (2014b). Water stress and recovery in the performance of two *Eucalyptus globulus* clones: physiological and biochemical profiles. *Physiol. Plant.* 150, 580–592. doi: 10.1111/ppl.12110
- Correia, B., Valledor, L., Hancock, R. D., Jesus, C., Amaral, J., Meijón, M., et al. (2016a). Depicting how *Eucalyptus globulus* survives drought: involvement of redox and DNA methylation events. *Funct. Plant Biol.* 43, 838–850. doi: 10.1071/FP16064
- Correia, B., Valledor, L., Hancock, R. D., Renaut, J., Pascual, J., Soares, A. M., et al. (2016b). Integrated proteomics and metabolomics to unlock global and clonal responses of *Eucalyptus globulus* recovery from water deficit. *Metabolomics* 12:141. doi: 10.1007/s11306-016-1088-4
- Das, K., and Roychoudhury, A. (2014). Reactive oxygen species (ROS) and response of antioxidants as ROS-scavengers during environmental stress in plants. *Front. Environ. Sci.* 2:53. doi: 10.3389/fenvs.2014.00053
- Dixon, R. A., and Paiva, N. L. (1995). Stress-induced phenylpropanoid metabolism. *Plant Cell* 7, 1085–1097. doi: 10.1105/tpc.7.7.1085
- Dobra, J., Motyka, V., Dobrev, P., Malbeck, J., Prasil, I. T., Haisel, D., et al. (2010). Comparison of hormonal responses to heat, drought and combined stress in tobacco plants with elevated proline content. *J. Plant Physiol.* 167, 1360–1370. doi: 10.1016/j.jplph.2010.05.013
- Dodd, A. N., Salathia, N., Hall, A., Kévei, E., Tóth, R., Nagy, F., et al. (2005). Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage. *Science* 309, 630–633. doi: 10.1126/science.1115581
- Duan, H., Duursma, R. A., Huang, G., Smith, R. A., Choat, B., O'grady, A. P., et al. (2014). Elevated [CO<sub>2</sub>] does not ameliorate the negative effects of elevated temperature on drought-induced mortality in *Eucalyptus radiata* seedlings. *Plant Cell Environ.* 37, 1598–1613. doi: 10.1111/pce.12260
- Durgbanshi, A., Arbona, V., Pozo, O., Miersch, O., Sancho, J. V., and Gómez-Cadenas, A. (2005). Simultaneous determination of multiple phytohormones in plant extracts by liquid chromatography-electrospray tandem mass spectrometry. *J. Agric. Food Chem.* 53, 8437–8442. doi: 10.1021/jf050884b
- Escandón, M., Cañal, M. J., Pascual, J., Pinto, G., Correia, B., Amaral, J., et al. (2016). Integrated physiological and hormonal profile of heat-induced thermotolerance in *Pinus radiata*. *Tree Physiol.* 36, 63–77. doi: 10.1093/treephys/tpv127
- Falcone, D. L., Ogas, J. P., and Somerville, C. R. (2004). Regulation of membrane fatty acid composition by temperature in mutants of *Arabidopsis* with alterations in membrane lipid composition. *BMC Plant Biol.* 4:17. doi: 10.1186/1471-2229-4-17
- Foito, A., Byrne, S. L., Hackett, C. A., Hancock, R. D., Stewart, D., and Barth, S. (2013). Short-term response in leaf metabolism of perennial ryegrass (*Lolium perenne*) to alterations in nitrogen supply. *Metabolomics* 9, 145–156. doi: 10.1007/s11306-012-0435-3
- Hancock, R. D., Morris, W. L., Ducreux, L. J., Morris, J. A., Usman, M., Verrall, S. R., et al. (2014). Physiological, biochemical and molecular responses of the potato (*Solanum tuberosum* L.) plant to moderately elevated temperature. *Plant Cell Environ.* 37, 439–450. doi: 10.1111/pce.12168
- Hochberg, U., Degu, A., Toubiana, D., Gendler, T., Nikoloski, Z., Rachmilevitch, S., et al. (2013). Metabolite profiling and network analysis reveal coordinated changes in grapevine water stress response. *BMC Plant Biol.* 13:184. doi: 10.1186/1471-2229-13-184

- Hodges, D. M., Delong, J. M., Forney, C. F., and Prange, R. K. (1999). Improving the thiobarbituric acid-reactive-substances assay for estimating lipid peroxidation in plant tissues containing anthocyanin and other interfering compounds. *Planta* 207, 604–611. doi: 10.1007/s004250050524
- Horváth, E., Szalai, G., and Janda, T. (2007). Induction of abiotic stress tolerance by salicylic acid signaling. *J. Plant Growth Regul.* 26, 290–300. doi: 10.1007/s00344-007-9017-4
- Larkindale, J., Hall, J. D., Knight, M. R., and Vierling, E. (2005). Heat stress phenotypes of Arabidopsis mutants implicate multiple signaling pathways in the acquisition of thermotolerance. *Plant Physiol.* 138, 882–897. doi: 10.1104/pp.105.062257
- Lê Cao, K.-A., Rohart, F., Gonzalez, I., and Dejean, S. (2016). *mixOmics: Omics Data Integration Project. R Package*. Available at: <https://CRAN.R-project.org/package=mixOmics>
- Lindner, M., Maroschek, M., Netherer, S., Kremer, A., Barbati, A., Garcia-Gonzalo, J., et al. (2010). Climate change impacts, adaptive capacity, and vulnerability of European forest ecosystems. *For. Ecol. Manage.* 259, 698–709. doi: 10.1016/j.foreco.2009.09.023
- Lyon, D., Castillejo, M. A., Mehmeti-Tershani, V., Staudinger, C., Kleemaier, C., and Wienkoop, S. (2016). Drought and Recovery: Independently regulated processes highlighting the importance of protein turnover dynamics and translational regulation in *Medicago truncatula*. *Mol. Cell. Proteomics* 15, 1921–1937. doi: 10.1074/mcp.M115.049205
- Mitchell, P. J., Battaglia, M., and Pinkard, E. A. (2013). Counting the costs of multiple stressors: is the whole greater than the sum of the parts. *Tree Physiol.* 33, 447–450. doi: 10.1093/treephys/tp031
- Mittler, R. (2006). Abiotic stress, the field environment and stress combination. *Trends Plant Sci.* 11, 15–19. doi: 10.1016/j.tplants.2005.11.002
- Mittler, R., and Blumwald, E. (2010). Genetic engineering for modern agriculture: challenges and perspectives. *Annu. Rev. Plant Biol.* 61, 443–462. doi: 10.1146/annurev-arplant-042809-112116
- Obata, T., Witt, S., Lisek, J., Palacios-Rojas, N., Florez-Sarasa, I., Yousfi, S., et al. (2015). Metabolite profiles of maize leaves in drought, heat, and combined stress field trials reveal the relationship between metabolism and grain yield. *Plant Physiol.* 169, 2665–2683. doi: 10.1104/pp.15.01164
- Osaki, M., Shinano, T., and Tadano, T. (1991). Redistribution of carbon and nitrogen compounds from the shoot to the harvesting organs during maturation in field crops. *Soil Sci. Plant Nutr.* 37, 117–128. doi: 10.1080/00380768.1991.10415017
- Pandey, P., Ramegowda, V., and Senthil-Kumar, M. (2015). Shared and unique responses of plants to multiple individual stresses and stress combinations: physiological and molecular mechanisms. *Front. Plant Sci.* 6:723. doi: 10.3389/fpls.2015.00723
- Queval, G., and Noctor, G. (2007). A plate reader method for the measurement of NAD, NADP, glutathione, and ascorbate in tissue extracts: application to redox profiling during *Arabidopsis* rosette development. *Anal. Biochem.* 363, 58–69. doi: 10.1016/j.ab.2007.01.005
- R Core Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Rennenberg, H., Loreto, F., Polle, A., Brilli, F., Fares, S., Beniwal, R., et al. (2006). Physiological responses of forest trees to heat and drought. *Plant Biol.* 8, 556–571. doi: 10.1055/s-2006-924084
- Rivero, R. M., Mestre, T. C., Mittler, R., Rubio, F., Garcia-Sanchez, F., and Martinez, V. (2014). The combined effect of salinity and heat reveals a specific physiological, biochemical and molecular response in tomato plants. *Plant Cell Environ.* 37, 1059–1073. doi: 10.1111/pce.12199
- Rizhsky, L., Liang, H., and Mittler, R. (2002). The combined effect of drought stress and heat shock on gene expression in tobacco. *Plant Physiol.* 130, 1143–1151. doi: 10.1104/pp.006858
- Sharkey, T. D. (2005). Effects of moderate heat stress on photosynthesis: importance of thylakoid reactions, rubisco deactivation, reactive oxygen species, and thermotolerance provided by isoprene. *Plant Cell Environ.* 28, 269–277. doi: 10.1111/j.1365-3040.2005.01324.x
- Silva, E. N., Ferreira-Silva, S. L., De Vasconcelos Fontenele, A., Ribeiro, R. V., Viégas, R. A., and Silveira, J. A. G. (2010). Photosynthetic changes and protective mechanisms against oxidative damage subjected to isolated and combined drought and heat stresses in *Jatropha curcas* plants. *J. Plant Physiol.* 167, 1157–1164. doi: 10.1016/j.jplph.2010.03.005
- Sims, D. A., and Gamon, J. A. (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sens. Environ.* 81, 337–354. doi: 10.1016/s0034-4257(02)00010-x
- Smirnoff, N., and Cumbe, Q. J. (1989). Hydroxyl radical scavenging activity of compatible solutes. *Phytochemistry* 28, 1057–1060. doi: 10.1016/0031-9422(89)80182-7
- Sweetlove, L. J., Beard, K. F., Nunes-Nesi, A., Fernie, A. R., and Ratcliffe, R. G. (2010). Not just a circle: flux modes in the plant TCA cycle. *Trends Plant Sci.* 15, 462–470. doi: 10.1016/j.tplants.2010.05.006
- Tohge, T., Scossa, F., and Fernie, A. R. (2015). Integrative approaches to enhance understanding of plant metabolic pathway structure and regulation. *Plant Physiol.* 169, 1499–1511. doi: 10.1104/pp.15.01006
- Valladares, F., and Pearcy, R. (1997). Interactions between water stress, sun-shade acclimation, heat tolerance and photoinhibition in the sclerophyll *Heteromeles arbutifolia*. *Plant Cell Environ.* 20, 25–36. doi: 10.1046/j.1365-3040.1997.d01-8.x
- Wang, W., Vinocur, B., and Altman, A. (2003). Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. *Planta* 218, 1–14. doi: 10.1007/s00425-003-1105-5
- Wang, Y., Zhang, H., Hou, P., Su, X., Zhao, P., Zhao, H., et al. (2014). Foliar-applied salicylic acid alleviates heat and high light stress induced photoinhibition in wheat (*Triticum aestivum*) during the grain filling stage by modulating the psbA gene transcription and antioxidant defense. *Plant Growth Regul.* 73, 289–297. doi: 10.1007/s10725-014-9889-9
- Warren, C. R., Aranda, I., and Cano, F. J. (2011). Responses to water stress of gas exchange and metabolites in *Eucalyptus* and *Acacia* spp. *Plant Cell Environ.* 34, 1609–1629. doi: 10.1111/j.1365-3040.2011.02357.x
- Xu, Y.-H., Liao, Y.-C., Zhang, Z., Liu, J., Sun, P.-W., Gao, Z.-H., et al. (2016). Jasmonic acid is a crucial signal transducer in heat shock induced sesquiterpene formation in *Aquilaria sinensis*. *Sci. Rep.* 6:21843. doi: 10.1038/srep21843
- Xu, Z., Jiang, Y., and Zhou, G. (2015). Response and adaptation of photosynthesis, respiration, and antioxidant systems to elevated CO<sub>2</sub> with environmental stress in plants. *Front. Plant Sci.* 6:701. doi: 10.3389/fpls.2015.00701
- Zandalinas, S. I., Mittler, R., Balfagón, D., Arbona, V., and Gómez-Cadenas, A. (2017). Plant adaptations to the combination of drought and high temperatures. *Physiol. Plant.* 162, 2–12. doi: 10.1111/ppl.12540
- Zandalinas, S. I., Rivero, R. M., Martínez, V., Gómez-Cadenas, A., and Arbona, V. (2016). Tolerance of citrus plants to the combination of high temperatures and drought is associated to the increase in transpiration modulated by a reduction in abscisic acid levels. *BMC Plant Biol.* 16:105. doi: 10.1186/s12870-016-0791-7

**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 Correia, Hancock, Amaral, Gomez-Cadenas, Valledor and Pinto. 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.



# Molecular Profiling of Pierce's Disease Outlines the Response Circuitry of *Vitis vinifera* to *Xylella fastidiosa* Infection

Paulo A. Zaini<sup>1†</sup>, Rafael Nascimento<sup>2†</sup>, Hossein Gouran<sup>1</sup>, Dario Cantu<sup>3</sup>, Sandeep Chakraborty<sup>1</sup>, My Phu<sup>1</sup>, Luiz R. Goulart<sup>2</sup> and Abhaya M. Dandekar<sup>1\*</sup>

<sup>1</sup> Department of Plant Sciences, University of California, Davis, Davis, CA, United States, <sup>2</sup> Institute of Genetics and Biochemistry, Federal University of Uberlândia, Uberlândia, Brazil, <sup>3</sup> Department of Viticulture and Enology, University of California, Davis, Davis, CA, United States

## OPEN ACCESS

### Edited by:

Glória Catarina Pinto,  
University of Aveiro, Portugal

### Reviewed by:

Jorge Martín-García,  
University of Valladolid, Spain  
Mónica Meijón,  
Universidad de Oviedo, Spain

### \*Correspondence:

Abhaya M. Dandekar  
amdandekar@ucdavis.edu

<sup>†</sup> These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Plant Systems and Synthetic Biology,  
a section of the journal  
Frontiers in Plant Science

**Received:** 17 February 2018

**Accepted:** 18 May 2018

**Published:** 08 June 2018

### Citation:

Zaini PA, Nascimento R, Gouran H,  
Cantu D, Chakraborty S, Phu M,  
Goulart LR and Dandekar AM (2018)  
Molecular Profiling of Pierce's Disease  
Outlines the Response Circuitry  
of *Vitis vinifera* to *Xylella fastidiosa*  
Infection. *Front. Plant Sci.* 9:771.  
doi: 10.3389/fpls.2018.00771

Pierce's disease is a major threat to grapevines caused by the bacterium *Xylella fastidiosa*. Although devoid of a type 3 secretion system commonly employed by bacterial pathogens to deliver effectors inside host cells, this pathogen is able to influence host parenchymal cells from the xylem lumen by secreting a battery of hydrolytic enzymes. Defining the cellular and biochemical changes induced during disease can foster the development of novel therapeutic strategies aimed at reducing the pathogen fitness and increasing plant health. To this end, we investigated the transcriptional, proteomic, and metabolomic responses of diseased *Vitis vinifera* compared to healthy plants. We found that several antioxidant strategies were induced, including the accumulation of gamma-aminobutyric acid (GABA) and polyamine metabolism, as well as iron and copper chelation, but these were insufficient to protect the plant from chronic oxidative stress and disease symptom development. Notable upregulation of phytoalexins, pathogenesis-related proteins, and various aromatic acid metabolites was part of the host responses observed. Moreover, upregulation of various cell wall modification enzymes followed the proliferation of the pathogen within xylem vessels, consistent with the intensive thickening of vessels' secondary walls observed by magnetic resonance imaging. By interpreting the molecular profile changes taking place in symptomatic tissues, we report a set of molecular markers that can be further explored to aid in disease detection, breeding for resistance, and developing therapeutics.

**Keywords:** defense response, Xanthomonadaceae, plant-bacteria interaction, vascular pathogen, transcriptome, proteome, metabolome

## INTRODUCTION

Plants have evolved complex responses to adapt to both biotic and abiotic environmental stresses, with an increase in the production of reactive oxygen species (ROS) as a key mechanism common to several types of stress conditions (Mittler, 2017). Genome sequencing and other high-throughput approaches have greatly advanced the identification of specific responses, enabling the classification of hundreds of genes responsive to particular stresses. In *Vitis vinifera* for example,

different abiotic and biotic stressors have been studied (Lin et al., 2007; Choi et al., 2013; Abou-Mansour et al., 2015; Burger and Maree, 2015; Czempler et al., 2015; Dadakova et al., 2015; Kelloniemi et al., 2015; Savoi et al., 2016), which have provided important insights on the molecular aspects of Pierce's disease (PD) development following infection by the bacterium *Xylella fastidiosa*. The disease can be transmitted by infected plant material and tools, as well as by many species of xylem sap-feeding insects that vector the pathogen from plant to plant (Chatterjee et al., 2008). Detection and spread of different strains in the European continent in recent years have raised great concern, given its potential to colonize plant hosts already attacked in the American continent such as grapevines and oranges among others. It also poses the threat of causing new diseases like olive quick decline syndrome, which is annihilating olive groves in southern Italy and progressively spreading in the Mediterranean area (Martelli et al., 2016). Disease progression varies widely depending on environmental conditions and genotypes of pathogen and host scion and rootstocks, collectively contributing to different chemical microenvironments of scion sap (Wallis et al., 2013; Katam et al., 2015). Successful microbe proliferation and long distance movement across xylem elements are mediated by secreted virulence factors such as the polygalacturonase PglA, the lipase/esterase LesA, and protease PrtA that are able to modify xylem integrity by enzymatic activity (Agüero et al., 2005; Roper et al., 2007; Gouran et al., 2016; Nascimento et al., 2016), and together contribute to disease progression. Vascular occlusions caused by pathogen biofilm formation, host cell-wall thickening by callose deposition and lignification, as well as tylose formation all contribute to reduction of sap flow leading to water and nutrient limitation as symptoms progress (Chatterjee et al., 2008; Choi et al., 2013; Sun et al., 2013). Despite significant advances, however, the grapevine response circuitry to PD is still not thoroughly characterized. Here we report a systems biology approach to expand our understanding of the cellular and molecular changes during PD development in Thompson seedless grapevines. Our data highlighted major metabolites accumulated in PD and also revealed novel members of the pathogen-sensing and stress response network. This enabled us to select which genes within paralog groups play a more pronounced role in the defense response to PD and thus can be further explored as early disease markers or therapeutic targets in case of disease susceptibility genes. Since current mitigation strategies to control PD rely on intensive insecticide applications to prevent vectors from disseminating *X. fastidiosa* across grapevines, understanding disease susceptibility and the host molecular responses to infection can lead to improved resistance breeding and novel control approaches.

## MATERIALS AND METHODS

### *V. vinifera* Inoculation With *X. fastidiosa* and Preparation of Leaf Extracts

Controlled inoculations of 3-month-old clonally propagated grapevines (*V. vinifera* var. Thompson Seedless) were performed

according to (Dandekar et al., 2012). Briefly, plants were laid horizontally and 10  $\mu$ L of succinate-citrate buffer containing  $\sim 10^6$  cells of *X. fastidiosa* strain Temecula1 (NCBI Accession PRJNA285) was deposited on the cane  $\sim 10$  cm above soil level and punctured with a needle to allow for uptake of bacterial suspension into the xylem. Plants were kept in greenhouse conditions (20–25°C and watered daily) and leaf samples collected 12 weeks post inoculations. Preparation of grapevine leaf extracts was done as described in Nascimento et al. (2016) using three complete leaves excluding petioles from eight infected and eight control (inoculated with succinate-citrate buffer without *X. fastidiosa*) plants. Briefly leaves were flash frozen in liquid nitrogen, ground with mortar and pestle, and kept at  $-80^\circ\text{C}$  until use.

### Nuclear Magnetic Resonance Imaging

Nuclear magnetic resonance imaging ( $^1\text{H}$ -MRI) was done in an Avance 400 spectrometer equipped with Bruker DRX console microimaging accessory according to (Dandekar et al., 2012). Stem transverse sections of all non-infected and infected plants were collected between internodes located at the top (apical), middle, and bottom of the central stem (three cuts per plant) and subjected to MRI. **Figure 1C** shows representative results.

### RNA Extraction, Library Preparation, and Sequencing

RNA extraction from 1 g of ground leaf tissue was done with MasterPure Complete kit (Epicentre, IL) from five infected and five healthy (control) plants. Strand-specific RNA-seq libraries were generated by the UC Davis Genome Center DNA Technologies Core Facility from the ribo-depleted RNA samples using an Apollo 324 liquid handler (Wafergen, CA) and PrepX RNA library preparation kits (Wafergen, CA) following the instructions of the manufacturer. After a cleanup step using 1x volume of Ampure XP beads (Beckman Coulter, CA), the single-end RNA-seq libraries were PCR-amplified using Phusion High-Fidelity polymerase (NEB, MA) following standard procedures, cleaned up again using a 1x volume of Ampure XP beads, and then quantified by fluorometry (Qubit; LifeTechnologies, CA). Libraries were analyzed with a Bioanalyzer 2100 instrument (Agilent, CA) and then pooled in equimolar ratios according to the fluorometric measurements. The pooled libraries (from five infected plants and five non-infected plants) were quantified by qPCR with a Kapa Library Quant kit (Kapa, South Africa) and sequenced on one lane of an Illumina HighSeq 2500 (Illumina, CA).

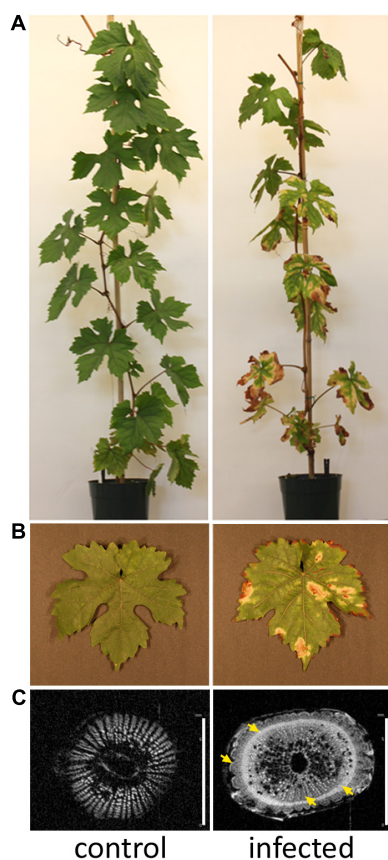
### Transcriptome Data Analysis and Validation

RNA-seq read quality and contamination were assessed with FastQC v.0.10.1<sup>1</sup>. Scythe v.0.991 and Sickle v.1.210<sup>2</sup> were used for Illumina adapter and quality-based trimming, respectively. Reads trimmed to less than 25 bases were discarded. *V. vinifera*

<sup>1</sup><http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

<sup>2</sup><https://github.com/ucdavis-bioinformatics>





**FIGURE 1 |** Grapevines developing Pierce's disease in controlled inoculations. **(A)** Twelve weeks after inoculation of *X. fastidiosa*, grapevines already displayed Pierce's disease symptoms as leaf scorching, match stick petioles from fallen leaves, and darkened patches on stems. **(B)** Detail of leaf showing scorching symptoms. **(C)** Nuclear magnetic resonance imaging of representative transversal cuts of stems exhibiting intensive secondary wall deposition in infected vines. Yellow arrows point to electron-denser material more abundant in infected samples. Scale bar is 6 mm for both images.

genome assembly version IGPP 12x and annotation data used in this analysis can be found at Genoscope (Jaillon et al., 2007) and Ensembl Gramene release 51 (Tello-Ruiz et al., 2016).

Reads were aligned to the *V. vinifera* genome using bowtie2 v.2.1.0 (Langmead and Salzberg, 2012). A reference transcriptome was generated from the NCBI files, using the gffread program within cufflinks v.2.1.1 (Roberts et al., 2011). BWA's short read aligner v.0.6.2 (Ardales et al., 2009) was then used to align the reads to the augmented transcriptome. Raw counts per gene were generated from the bwa alignments using sam2counts.py<sup>3</sup>. The raw counts from each of the five diseased and five control samples were statistically analyzed with EdgeR (Robinson et al., 2010) to produce tables of expression values, fold changes, and selection of differentially expressed (DE) transcripts, using  $p < 0.05$  or  $0.01$  from the fitted negative binomial generalized linear models and quasi-likelihood F-test (Supplementary Tables S2–S5). Enrichment of gene ontology

<sup>3</sup><https://github.com/ucdavis-bioinformatics>

terms of DE transcripts was performed with PANTHER using Bonferroni correction and  $p$ -value cutoff of 0.05 (Mi et al., 2016). Reverse transcription – quantitative polymerase chain reaction (RT-qPCR) was performed as detailed in Dandekar et al. (2012) as a means to verify the expression data obtained with RNA-seq, encompassing genes with distinct overall expression levels and ratios between infected and non-infected samples, as well as distinct functional categories. A two-tailed paired sample Student's  $t$ -test ( $\alpha = 0.05$ ) was performed with XLSTAT software on delta-Cq values to determine statistical significance of differences between infected and non-infected samples (three biological replicas assayed in duplicate each). One and two asterisks indicate, respectively,  $p$ -values  $< 0.05$  and  $0.01$ . Oligonucleotide primers used in RT-qPCR are listed in Supplementary Table S6. Transcriptome data have also been deposited in NCBI SRA under BioProject accession number PRJNA390670.

## Proteome Analysis

Protein extraction from 500 mg of ground grapevine leaf preparations (three diseased samples vs. three healthy controls) was performed with P-PER plant protein extraction reagent (Thermo Scientific). Samples were reconstituted in phosphate buffer saline (PBS) and 300  $\mu$ g were precipitated with 4x volume of precipitation reagent (CalBiochem) according to manufacturer's instructions. Precipitated samples were reconstituted in 100  $\mu$ l of 6 M urea + 5 mM 1,4-dithiothreitol (DTT) and incubated at 37°C for 30 min; 15 mM iodoacetamide (IAA) was added and incubated at room temperature for 30 min. IAA was then quenched with 30 mM DTT and incubated for 10 min. Lys-C/trypsin was added to a 1:25 enzyme:protein ratio and incubated at 37°C for 4 h; 50 mM ammonium bicarbonate was added to dilute urea and activate trypsin and digestion occurred overnight at 37°C. Digested peptides were then de-salted using Aspire RP30 Desalting Tips (Thermo Scientific) and resuspended in loading buffer.

The digested peptides were analyzed using a QExactive mass spectrometer (Thermo Fisher Scientific) coupled with an Easy-LC (Thermo Fisher Scientific) and a nanospray ionization source. The peptides were loaded onto a trap (100 micron, C18 100 Å 5U) and desalted online before separation using a reverse phased column (75 micron, C18 200 Å 3U). The gradient duration for separation of peptides was 60 min using 0.1% formic acid and 100% acetonitrile for solvents A and B, respectively. Data were acquired using a data-dependent ms/ms method, which had a full scan range of 300–1,600 Da and a resolution of 70,000. The ms/ms method's resolution was 17,500 and an isolation width of 2  $m/z$  with normalized collision energy of 27. The nanospray source was operated using 2.2 kV spray voltage and a heated transfer capillary temperature of 250°C. Raw data were analyzed using X!Tandem (Fenyo and Beavis, 2003) and visualized using Scaffold Proteome Software (version 4.4.1). A protein was considered identified when at least two peptides were mapped to it with  $>99\%$  confidence threshold. Proteins with differential abundance between infected and non-infected samples were chosen by  $p < 0.05$  obtained from the Mann–Whitney test. Samples were searched against Uniprot databases appended with the cRAP

database, which contains common laboratory contaminants. Reverse decoy databases were also applied to the database prior to the X!Tandem searches. Raw and differential analysis data are presented in Supplementary Tables S7, S8. The proteome procedure was performed at the UC Davis Proteomics Core.

## Immunodetection of Proteins

Anti-ferritin HRP-conjugated polyclonal antibody was generated in rabbit by injecting synthetic peptides corresponding to structural epitopes (GenScript, NJ). Antibody was diluted in PBS-M 1% (PBS plus 1% non-fat dried milk) at a 1:500 dilution. Blocking and washing used PBS-M 5% (PBS plus 5% non-fat dried milk) and PBS-T 0.1% (PBS plus 0.1% Tween 20), respectively, and blots were developed using ECL Plus Western Blotting Detection Reagents (GE Life Sciences, United States) and visualized using a ChemiDoc-It TS2 (BioRad, CA) imaging instrument.

## Metabolome Analysis

Metabolomic analysis was performed at the NIH West Coast Metabolomics Center, UC Davis. Sample preparation and gas chromatography coupled to time-of-flight mass spectrometry (GC-TOF/MS) followed the protocols in Sana et al. (2010), including data filtering, BinBase assignment, and statistics. Eight samples from diseased plants vs. eight control healthy samples were compared. Metabolite extraction was done on 100 mg of ground leaf tissue extracted for 20 min at  $-20^{\circ}\text{C}$  in pre-cooled 2:3:3 v/v/v solvent mixture of water/acetonitrile/isopropanol and centrifuged 16,000 *g* for 3 min. The liquid phase supernatant was used for GC-TOF/MS performed on an Agilent 6890 gas chromatograph (Santa Clara, CA, United States) controlled by the Leco ChromaTOF software vs. 2.32 (St. Joseph, MI, United States). A 30 m long, 0.25 mm internal diameter rtx5Sil-MS column with 0.25  $\mu\text{m}$  5% diphenyl/95% dimethyl polysiloxane film and additional 10 m integrated guard column was used (Restek, Bellefonte, PA, United States). Absolute spectra intensities were processed by a filtering algorithm implemented in the metabolomics BinBase database (Fiehn et al., 2005). Quantification was reported as peak height using the unique ion as default, unless a different quantification ion was manually set in the BinBase administration software Bellerophon. Metabolites were unambiguously assigned by the BinBase identifier numbers, using retention index and mass spectrum as the two most important identification criteria. Additional confidence criteria were given by mass spectral metadata, using the combination of unique ions, apex ions, peak purity, and signal/noise ratios as given in data preprocessing. All database entries in BinBase were matched against the Fiehn mass spectral library of 1,200 authentic metabolite spectra and the NIST05 commercial library (Kind et al., 2009). BinBase entries were named manually by both matching mass spectra and retention index. Statistical evaluation was performed using univariate Student's *t*-test for independent pairs of groups. Values of  $p < 0.05$  were considered statistically significant. Raw data and differential analysis are presented in Supplementary Tables S9, S10. Multivariate statistical analysis was performed with XLSTAT software by considering the Pearson correlation

among the samples and plotted using Circos (Krzywinski et al., 2009).

## RESULTS

### Multiomic Analysis of Pierce's Disease Delineates the Pathogen Perception and Host Response Circuitry

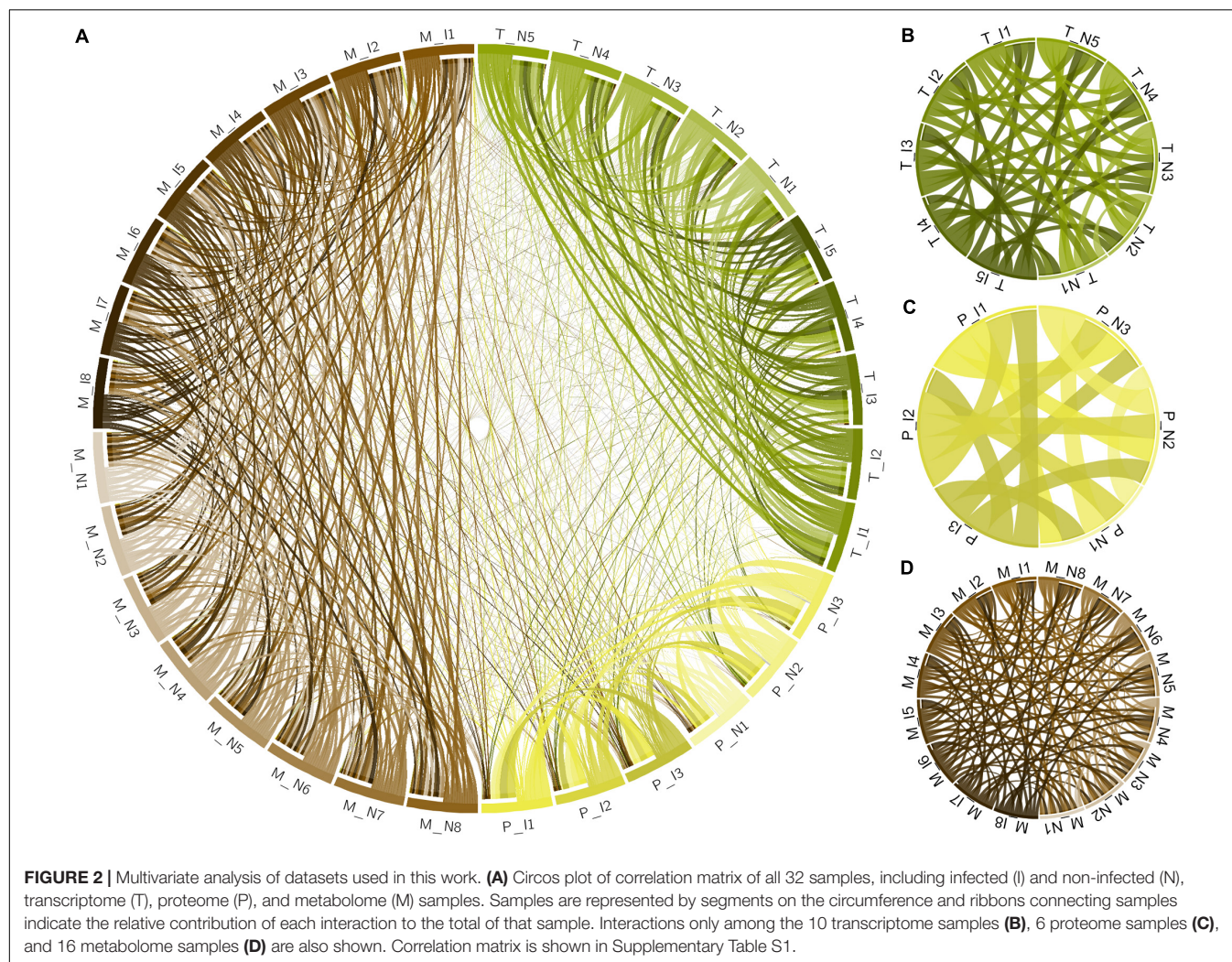
*Xylella fastidiosa*-infected grapevines under controlled greenhouse conditions started to show initial PD symptoms on leaves and stems  $\sim 8$  weeks post inoculation. These were initially limited to scorching of leaf blades near margins, progressing inward toward petioles. By the time, our samples were collected 12 weeks post inoculation, brownish patches were already visible on stems, and some leaves had already dropped off leaving typical "matchstick" petioles and an overall dehydrated appearance (Figures 1A,B). Transversal cuts of stems near the apical meristem showed that besides the brownish patches on canes, intensive alterations were also occurring inside the stems, marked by an increase in secondary cell wall deposition and thickening on infected vines (Figure 1C). Under these conditions, the leaf samples were collected and processed for RNA-seq deep sequencing, isobaric labeling proteome mass spectrometry, and GS-TOF/MS metabolite profiling. With all three methodologies, we were able to identify analytes over- and under-represented in symptomatic vines, both of known and unknown functions (Table 1). Multivariate analysis was used to determine the correlation among the datasets used in this work (Figure 2), which shows as expected that experimental technique (transcriptome, proteome, or metabolome) is a stronger determinant of higher correlation than experimental group (infected or non-infected). Considering only transcripts with low variability among replicas ( $p < 0.05$ ), we found only 16 genes that we also detected as DE in the proteome dataset. These encompass ATP synthase subunit alpha (VIT\_00s0733g00010), ADP/ATP carrier (VIT\_08s0007g02450), HtpG chaperone family protein (VIT\_02s0025g04340), dehydrin (VIT\_04s0023g02480),

**TABLE 1 |** Summary of multiomics data from grapevine leaves with Pierce's disease.

Transcriptome	Increased <sup>a</sup>	Decreased	Total
CDS with putative function	854	238	19140
CDS without putative function	123	25	4907
Total detected	977	263	23991
Proteome			
Proteins with putative function	20	15	98
Proteins without putative function	69	23	263
Total detected	89	38	361
Metabolome			
Identified metabolites	14	4	121
Unidentified metabolites	29	3	230
Total detected	43	7	351

<sup>a</sup>Increased and decreased abundance in diseased leaves compared to healthy controls.

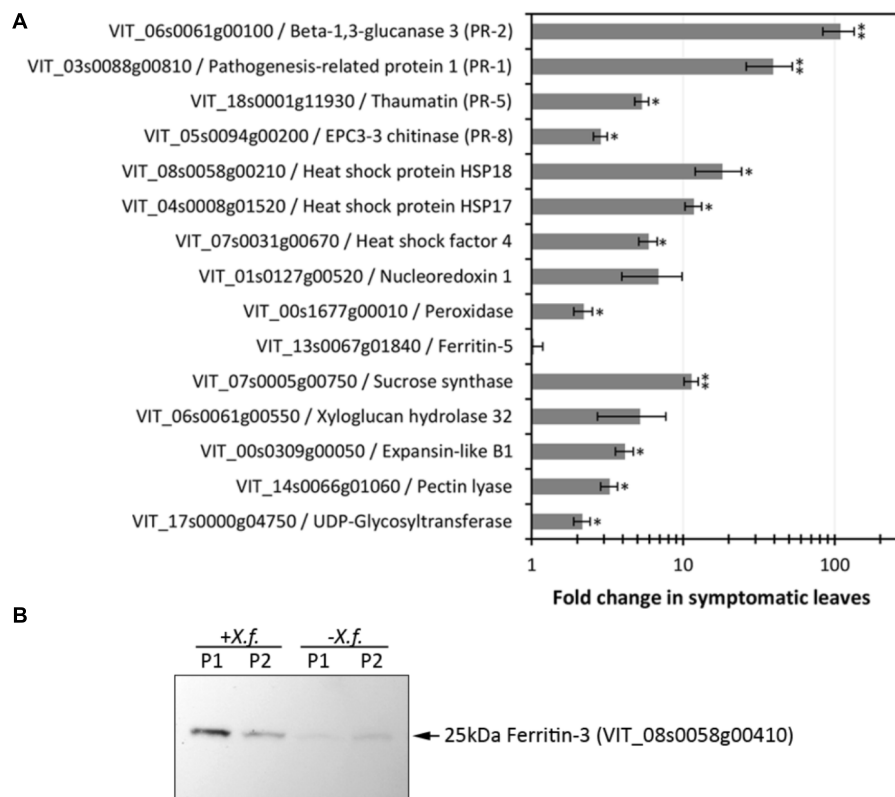




lipid-transfer protein (VIT\_05s0020g03750), beta 1,3-glucanase 3 (VIT\_06s0061g00120), alpha-beta hydrolase (VIT\_07s0005g01240), ferritin 3 (VIT\_08s0058g00410), thioredoxin superfamily protein (VIT\_12s0028g03010), chitinase-18 (VIT\_14s0066g00610), subtilisin-like proteins (VIT\_15s0048g01180, VIT\_18s0001g14870), PHB domain-containing membrane-associated protein (VIT\_16s0100g00090), MYB4 (VIT\_17s0000g04750), glycine-rich RNA-binding protein (VIT\_18s0001g11930), and a Clp protease (VIT\_19s0014g03160). Although these overlapping results from the transcriptome and proteome datasets illustrate fragments of the host response to infection, many other details were captured in each dataset and will be presented next. A complete dataset of detected metabolites, including their relative levels compared to those in healthy vines, are available as Supplementary Materials, and grouped by transcriptomic, proteomic, and metabolomic data (Supplementary Tables S1–S11). Among the methods employed in this work, our RNA-seq dataset is the most comprehensive and revealed expression for 23,991 unique protein coding sequences (CDS), representing 80% of the 29,971 CDS predicted in the CRIBI V1 grape genome annotation. To

further validate the transcriptome data, we performed RT-qPCR on 15 CDS from *X. fastidiosa*-infected leaf extracts encompassing a wide range of response intensities (**Figure 3A**) and western blot of ferritin (**Figure 3B**). These additional techniques also show that while some ferritins respond to disease, other paralogs do not, as previously suggested by the transcriptome and proteome data. More details on the ferritins will be presented ahead.

Among the 1,240 DE CDS, 47 were detected only in infected tissues including laccases, polygalacturonase, pectin lyase, and fasciclin-like arabinogalactan protein (FLA) genes mainly involved in cell-wall remodeling and lignification (Wang et al., 2015). A gene ontology analysis of all DE CDS shows enriched terms describing biological processes and molecular functions of these and other DE genes, as listed in **Table 2**. Clustering of transcripts by sequence similarity also reveals enriched functions in the transcriptome data, such as chalcone and stilbene synthases, laccases, and other cell wall remodeling enzymes (Supplementary Table S5). The systems approach was further extended to MapMan functional analysis (Thimm et al., 2004), revealing various aspects of the metabolic shift accompanying PD development, from perception of the pathogen



**FIGURE 3 |** Evaluation of expression levels of selected CDS and proteins. **(A)** RT-qPCR analysis of CDS selected based on RNA-seq data encompassing various functions and modulation intensities in samples from diseased leaves. Fold changes compared to healthy control plants were calculated from three biological replicates assayed in triplicate each. One and two asterisks indicate, respectively,  $p$ -values  $< 0.05$  and  $0.01$ , according to two-tailed  $t$ -test for two paired samples with  $\alpha = 0.05$ . **(B)** Immunodetection of ferritin 3 in leaf samples (P1, P2: two plants per treatment) infected with *X. fastidiosa* (Xf).

to cell wall modification, onset of oxidative stress response, and polyphenol metabolism to counteract chronic oxidative damage (Figure 4). Although convenient to visualize fold-changes of specific genes of major metabolic functions affected by the disease, we also sought to identify robust disease markers by including the notion of expression level of each marker, as shown in Figure 5. Taken together, these parameters of relative abundance provide a wealth of information describing the molecular events that will be explored in the following sections.

Differentially expressed CDS related to extracellular signal perception including 22 leucine-rich repeat (LRR) receptor-like kinases (RLKs) were highlighted, with emphasis to VIT\_14s0128g00550 and VIT\_10s0003g01430. We used the genes encoding LRR-RLKs as an example of the power of genome-wide investigations to select the paralogous members within the functional family group that are relevant to PD (Figure 6A). Transcripts encoding five cysteine-rich RLKs also showed increased abundance (VIT\_00s0262g00120, for example), indicating a multitude of input signal sources. CDS encoding an array of ankyrin-repeat proteins were also strongly modulated, especially VIT\_13s0106g00200 positively and VIT\_05s0165g00260 negatively. Signal transduction cascade MAPKs were also upregulated (VIT\_06s0004g06850, VIT\_07s0031g00530, and

VIT\_12s0059g00870) plus many other cytosolic kinases of various types. We also detected WRKY9 (VIT\_12s0055g00340), MYB108 (VIT\_05s0077g00500), GRAS (VIT\_13s0019g01700), and an AGAMOUS-like (VIT\_17s0000g01230) transcription factors (TFs) as the most responsive to PD, among many others modulated less intensively (Supplementary Table S4). Another interesting group of DE transcripts encode members of four of the seven families of nodulin-like proteins, including seven MtN21/UMAMIT-like (VIT\_04s0044g00450), five MtN3/SWEET-like (VIT\_11s0016g04920), two early nodulin-like (VIT\_14s0066g01420), and three vacuolar iron transporter/nodulin-like (VIT\_00s0267g00030).

## Pathogenesis-Related Proteins and Antimicrobial Compounds

Among the responses to *X. fastidiosa* infection, we found accumulation of transcripts encoding pathogenesis-related (PR) proteins, which constitute a complex repertoire of defense strategies aimed at inhibiting pathogen proliferation (Sels et al., 2008). These encompassed several  $\beta$ -1,3-glucanases (for example, VIT\_05s0077g01150), class I, II, III, and V chitinases (PR-11, PR-4, PR-8, PR-3, with VIT\_14s0066g00610 most intensively modulated), thaumatin-like proteins (PR-5, such as



**TABLE 2 |** Summary of PANTHER statistical overrepresentation test results for enrichment of GO terms in differentially expressed genes.

GO term	Term description	Fold enrichment	p-value	Transcript abundance
<i>Biological process</i>				
GO:0009698	Phenylpropanoid metabolic process	5.85	1.15E-03	+
GO:0009813	Flavonoid biosynthetic process	5.23	1.59E-03	+
GO:0046942	Carboxylic acid transport	4.95	1.64E-02	+
GO:0016052	Carbohydrate catabolic process	3.56	3.41E-02	+
GO:0032787	Monocarboxylic acid metabolic process	2.91	1.60E-03	+
GO:0071554	Cell wall organization or biogenesis	2.68	3.53E-03	+
GO:0055114	Oxidation-reduction process	1.77	4.64E-02	+
GO:2000031	Regulation of salicylic acid mediated signaling pathway	50.42	1.30E-04	–
GO:0016042	Lipid catabolic process	7.47	8.05E-03	–
<i>Molecular function</i>				
GO:0035251	UDP-glucosyltransferase activity	4.73	3.68E-02	+
GO:0048037	Cofactor binding	3.23	2.25E-04	+
GO:0022857	Transmembrane transporter activity	2.16	8.22E-03	+
GO:0016491	Oxidoreductase activity	1.84	6.43E-03	+
GO:0043167	Ion binding	1.35	4.73E-02	+
GO:0052689	Carboxylic ester hydrolase activity	6.44	1.83E-03	–
GO:0016705	Incorporation or reduction of molecular oxygen	4.82	2.71E-02	–
<i>Cellular component</i>				
GO:0071944	Cell periphery	1.64	2.29E-05	+
GO:0005886	Plasma membrane	1.58	1.78E-03	+
GO:0016021	Integral component of membrane	1.45	3.54E-04	+
GO:0005576	Extracellular region	2.45	6.87E-05	–

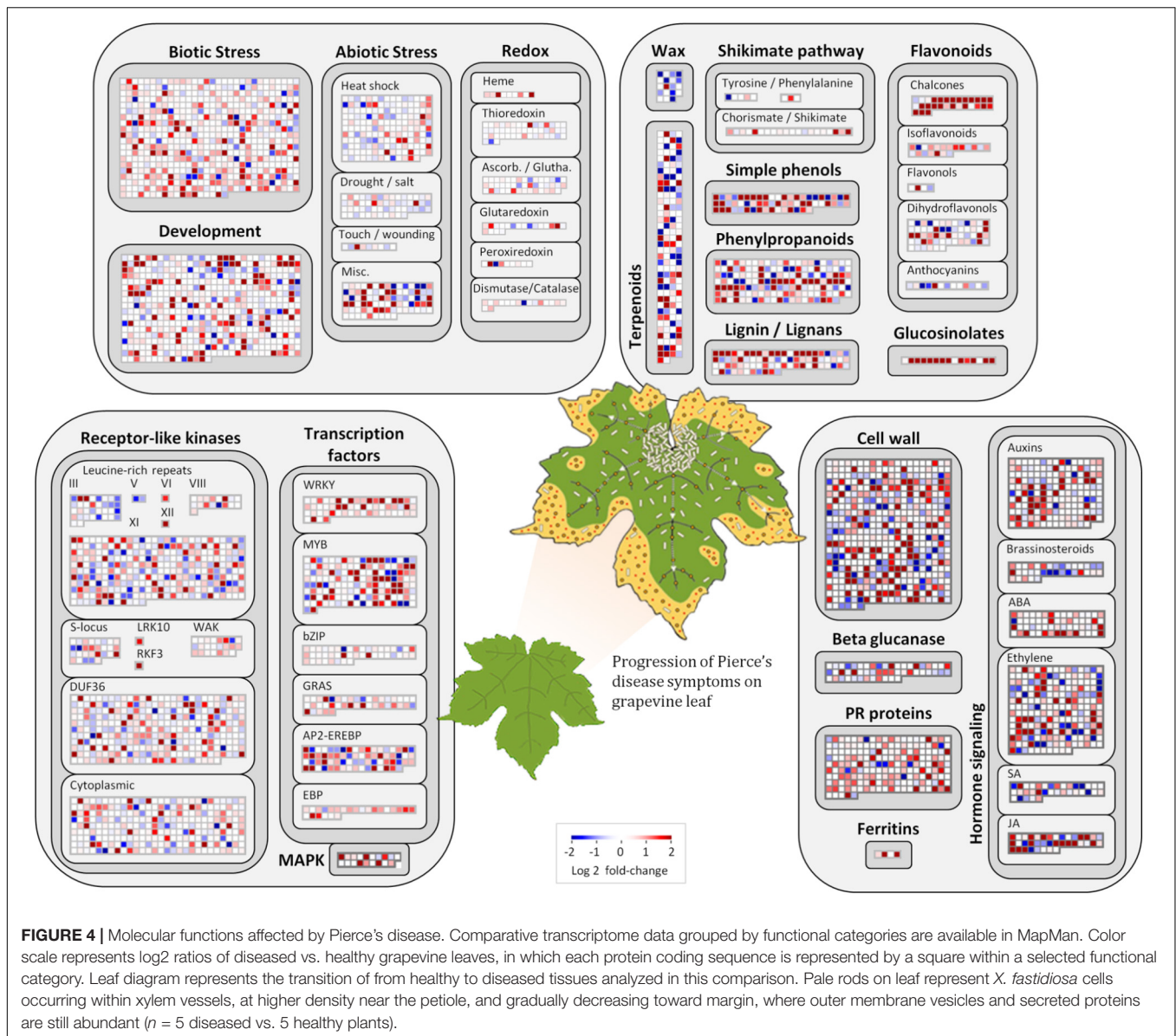
VIT\_18s0001g14480; **Figure 6B**), proteinase inhibitors (PR-6, VIT\_05s0020g05040), proteinases (PR-7, VIT\_07s0104g00180), peroxidases (PR-8, VIT\_14s0066g01670), ribonucleases (PR-9, VIT\_14s0060g01530), lipid transfer proteins (PR-14, VIT\_06s0004g08060 upregulated and VIT\_08s0058g01210 downregulated), and oxalate oxidase germins (PR-15 VIT\_09s0002g01340, and PR-16 VIT\_07s0005g02370). The proteome data also highlight the chitinase VIT\_14s0066g00610 as one of the most modulated in infected tissues, corroborating the transcriptome data. Among the PR proteins, 12 CDS encoding germin-like proteins from the RmlC-like cupins superfamily were upregulated, with VIT\_14s0128g00600 in greater intensity. Three Kunitz-type protease inhibitors were also strongly induced (VIT\_17s0119g00150, among others).

Specialized (secondary) metabolism was also strongly influenced by disease onset. Twenty-two upregulated CDS for chalcone and stilbene synthases, involved in phytoalexin production against microbes (Aoki et al., 2016), and plant defense against abiotic stress such as UV-radiation were highlighted in our dataset, with VIT\_16s0100g01040 most intensively, and others in the same genomic vicinity of chromosome 16. Terpene metabolism was also modified according to our data, with increased abundance of transcripts for terpene synthases VIT\_18s0001g04120 and VIT\_00s0692g00020, while reducing that of 10 others with VIT\_19s0014g04930 and VIT\_00s0271g00010 most intensively. Enrichment of compounds known to inhibit pathogen growth and biofilm formation was also detected in our metabolome data such as erythritol and 2-deoxyerythritol (Ghezelbash et al., 2012),

1,2-anhydro-myo-inositol and arbutin; these latter glycosidase and tyrosinase inhibitors, respectively (Falshaw et al., 2000; Supplementary Table S10 and **Figure 7**). The pathogen might also benefit from increased concentration of nutritional compounds such as fructose, tryptophan, and glutamine.

## Oxidative, Drought, and Osmolarity Stress-Related Responses

Iron sequestration was triggered by disease onset, with ferritin isoforms VIT\_08s0058g00410, VIT\_08s0058g00430, and VIT\_08s0058g00440, being strongly induced, along with aconitase (VIT\_12s0059g02150) and nicotianamine synthase 4 (VIT\_11s0052g01150), involved in phytosiderophore production (Mizuno et al., 2003). Interestingly, these three ferritins encoded in chromosome 8 were intensively upregulated, while another present on chromosome 6 (VIT\_06s0004g07160) was not selected based on expression ratio, but remained highly expressed also in healthy samples. Another ferritin on chromosome 13 (VIT\_13s0067g01840) had very low expression in grapevine leaves with or without disease. Three vacuolar iron transporters on the other hand were downregulated (most markedly VIT\_00s0267g00030). Modulation of other iron-associated redox proteins such as ferric reductase, iron superoxide dismutase, and ferredoxin were not significant. Besides limiting iron availability to microbes, storing free iron within ferritin nanocages reduces ROS production and oxidative damage (Mehterov et al., 2012). Like in the transcriptome data, ferritin-3 (VIT\_08s0058g00410) was intensively upregulated

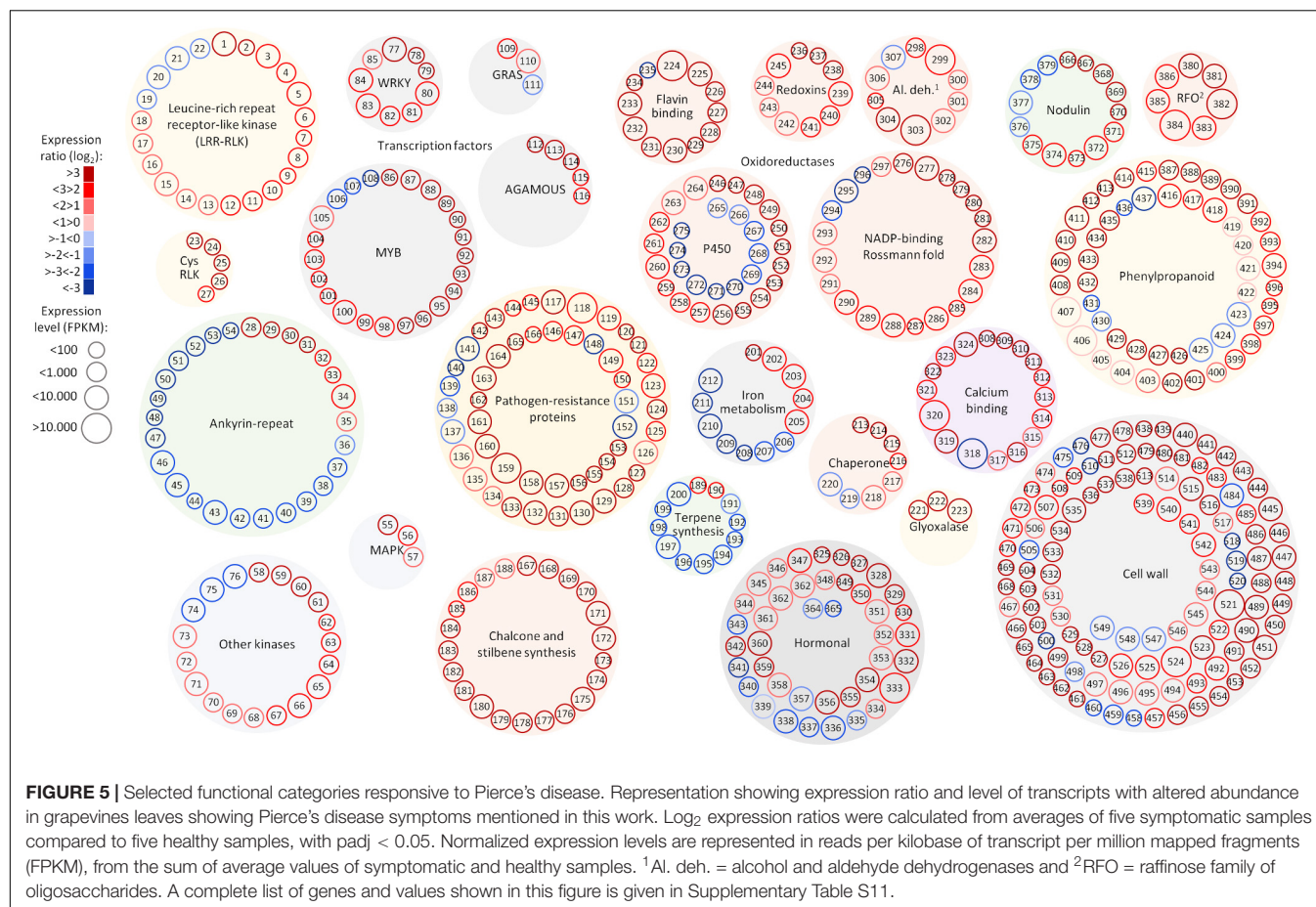


in the proteome data and further confirmed by western blot (**Figure 3B**), emerging as a robust feature of PD onset.

Other intense responses included upregulation of proteins and transcripts for low molecular weight heat shock proteins, glyoxalase I for methylglyoxal detoxification (for example, VIT\_09s0002g06430), and anthocyanidin reductases such as VIT\_15s0046g01150 involved in general stress tolerance (Xie et al., 2003; Hasanuzzaman et al., 2017). Anthocyanins evidenced in symptomatic leaves (**Figure 1B**) might come from the upregulated biosynthetic pathway encompassing phenylalanine ammonia-lyase (VIT\_11s0016g01640 and other copies), cinnamate-4-hydroxylase (VIT\_11s0078g00290), 4-coumarate CoA ligase (VIT\_06s0061g00450), chalcone synthase (VIT\_16s0100g01040 plus several others in this genomic vicinity), chalcone isomerase (VIT\_04s0008g02030), flavanone 3-hydroxylase (VIT\_18s0001g14310), leucoanthocyanidin

dioxygenase (VIT\_02s0025g04720), UDP-glucosyl transferase (VIT\_18s0001g12040 and others), and finally methyltransferases (VIT\_16s0100g00570 and others), possibly providing an array of different anthocyanins yet to be further characterized. Moreover, CDS for 21 NADP-binding Rossmann-fold enzymes with oxidoreductase activity were modulated, with VIT\_15s0046g01150 as the most prominent. Concurrent to this, our proteome data also showed photosynthesis components such as photosystem I and II, electron transport, carbon fixation, and ATP synthesis with decreased abundance, which could also contribute to a reduction of chlorophyll levels unmasking anthocyanins already present.

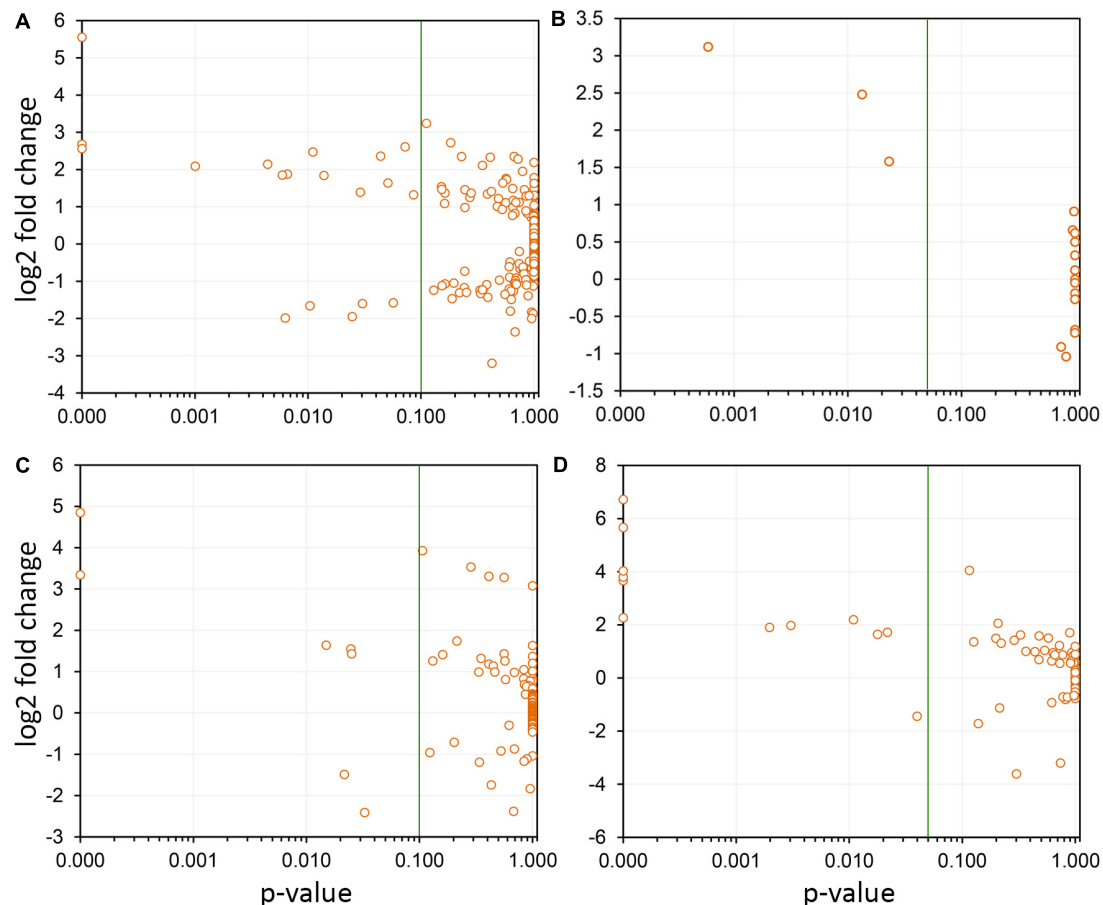
The metabolite with highest increased abundance detected on symptomatic leaves was gamma-aminobutyric acid (GABA), a known response for water, oxidative, and wounding



stress in *Arabidopsis thaliana* (Kinnersley and Turano, 2000; Coleman et al., 2001; Scholz et al., 2017). On the other hand, its precursor, glutamic acid, showed decreased abundance, supporting this intense metabolic flow (Supplementary Table S10 and Figure 7). These metabolome data are consistent with upregulation of two CDS encoding glutamate decarboxylase (GAD) in our transcriptome data (VIT\_01s0011g06600 and VIT\_01s0011g06610), which perform this enzymatic conversion. Our metabolome data also indicate several sugars upregulated in infected tissues, besides other metabolites with known antioxidant properties, including galactinol, catechin, cellobiose, and gentiobiose. These metabolites follow a trend observed in the transcriptome and proteome data in which several ROS-scavenging systems were also upregulated, such as three paralogs of galactinol synthase (VIT\_05s0077g00430 among others) and a raffinose hydrolase known as seed imbibition 2 (VIT\_08s0007g08310) which also helps to increase galactinol levels in response to pathogen attack and oxidative stress (Couee et al., 2006; Sengupta et al., 2015). Two other raffinose synthase family protein CDS (VIT\_11s0016g05770 and VIT\_07s0005g01680) were also highly upregulated possibly providing the substrate for galactinol synthesis. We also detected modulation of various cytochrome P450s with emphasis to VIT\_07s0129g00820, besides seven peroxidases (six with increased abundance

lead by VIT\_14s0066g01670 and VIT\_18s0001g06840). Other modulated enzymes involved in ROS turnover include five alcohol and five aldehyde dehydrogenases (VIT\_18s0001g15410 and VIT\_01s0026g00220, for example), glutathione S-transferases including VIT\_04s0079g00690 and Tau7-like VIT\_16s0039g01070, and quinone reductases encoded by VIT\_00s0271g00110 and VIT\_00s0274g00080, all of these playing a role in ROS turnover and byproduct detoxification. Interestingly, however, while some enzymes such as cysteine peroxiredoxin (VIT\_05s0020g00600) and thioredoxin (VIT\_12s0028g03010) accumulated both in the transcriptome and proteome data, other classical ROS-detoxifying enzymes such as superoxide dismutase, catalase and glutathione peroxidase did not. Other DE transcripts involved in ROS signaling also include CDS that bind calcium such calmodulins (VIT\_18s0122g00180 and five others), calcineurins (VIT\_17s0000g09480 with increased abundance and VIT\_17s0000g09470 with a sharp decrease), Ca<sup>2+</sup>-binding EF-hand family proteins (VIT\_12s0059g00340 and two others), as well as Ca<sup>2+</sup>-dependent lipid-binding (CaLB domain) family proteins. The EF-hand proteins are known to activate respiratory burst oxidase homologs (RBOH, VIT\_14s0060g02320 also upregulated in our data), capable of ROS production upon pathogen perception (Kadota et al., 2015). Moreover, three CDS encoding lactoylglutathione





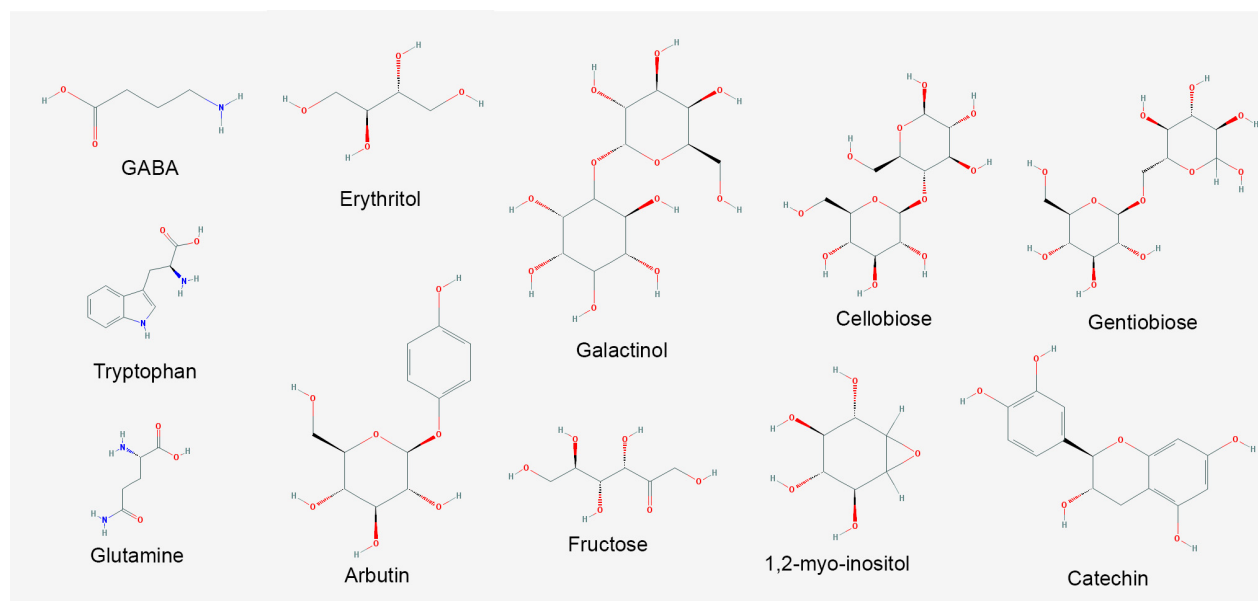
**FIGURE 6 |** Selection of paralogs responsive to Pierce's disease. Four examples are shown, in which the members of paralogous groups that are considered responsive to disease can be selected based on confidence levels (variability among replicates). In each panel, the green line indicates the  $p$ -value of 0.05 selection threshold used by us. **(A)** A total of 291 leucine-rich repeats receptor-like kinases were detected as expressed, and while only 19 were considered responsive to disease, 14 showed increased transcript abundance, and 5 reduced. **(B)** Seventeen thaumatin-like genes were detected as expressed, being three with increased transcript abundance. **(C)** From the 122 2-oxoglutarate oxygenase paralogs detected as expressed, seven were responsive to disease, with five showing increased transcript abundance. **(D)** From the 74 cellulose synthases detected as expressed, 12 were responsive to disease and of these only one shows reduced transcript abundance.

lyase/glyoxalase I (methylglyoxal detoxification), which are also calcium binding proteins are among the most induced in our dataset, particularly VIT\_11s0016g05010. Many of the responses to oxidative stress listed above are also detected in other kinds of stresses such as drought and osmolarity, exemplifying how PD results in various stresses to the plant host. All these responses interconnect resulting in the scorched leaves and other symptoms observed in diseased vines as shown in **Figure 1**. Another interesting connection between different stresses is major facilitator superfamily proteins for which eight CDS including VIT\_05s0020g02170 were upregulated and two were downregulated. Besides responding to salt stress, these sugar/H<sup>+</sup> symporters have been shown to be tightly correlated with programmed cell death (Norholm et al., 2006), as has flavin-dependent monooxygenases encoded by VIT\_11s0016g00570 and VIT\_07s0104g01260, also with increased abundance in diseased vines.

## Modulation of Hormone Biosynthesis and Signaling

Jasmonate biosynthesis was a strong hormonal response detected given the intensive positive regulation of seven 12-oxophytodienoate reductases including VIT\_18s0041g02020 and two allene oxide synthases (VIT\_03s0063g01820 and VIT\_18s0001g11630), known routes for jasmonic acid formation (Schaller et al., 2000). However, transcripts encoding downstream jasmonate O-methyltransferases (VIT\_04s0023g02230 among three other paralogs) and other DE CDS encoding enzymes of salicylic acid formation displayed a strong reduction in abundance. Despite the suggested inhibition of salicylic acid production, transcripts encoding the AGD2-like defense response protein 1 (*ALD1*, VIT\_18s0001g04630), accumulated to higher levels in diseased vines. On the catabolic side, the methyl salicylate esterase VIT\_00s0253g00140 was upregulated, possibly also contributing to lowered accumulation of salicylic acid. Gibberellin (GA) production also seems





**FIGURE 7 |** Major metabolites with increased abundance in *V. vinifera* affected by Pierce's disease. PubChem structures of metabolites with known antimicrobial and signaling functions accumulated in symptomatic grapevines ( $n = 8$  diseased vs. 8 healthy plants).

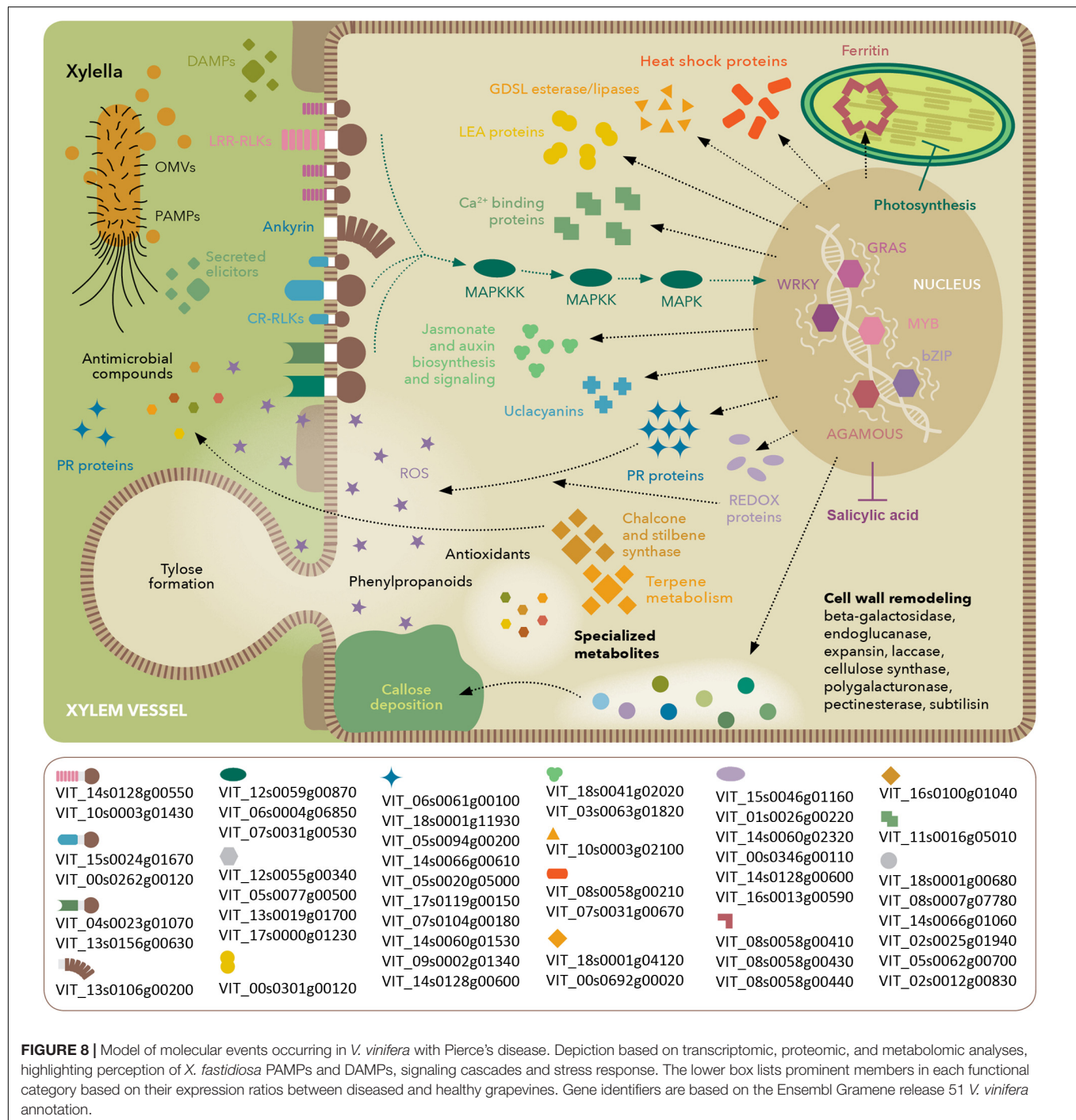
repressed as indicated by downregulation of 2 GA oxidases (VIT\_18s0001g01390 and VIT\_19s0177g00030), a MEP pathway member (VIT\_11s0052g01730), and a gibberellin-regulated protein (VIT\_08s0007g05860). Absciscic acid (ABA) biosynthesis epoxycarotenoid dioxygenase (VIT\_19s0093g00550), and ABA-responsive protein (VIT\_18s0001g10450), and two PP2C phosphatases (VIT\_06s0004g05460 and VIT\_16s0050g02680) were upregulated, another indication of the multiple stress physiology state of grapevines during PD. Although increased abundance of auxin biosynthesis enzymes was not detected, several auxin-responsive CDS were, including VIT\_12s0057g00420, three auxin efflux carriers VIT\_04s0044g01860, VIT\_05s0062g01120, and VIT\_04s0044g01880, auxin-induced protein VIT\_05s0049g01970, auxin-responsive of the GH3 type (VIT\_19s0014g04690, among three others), and TFs (VIT\_11s0016g04490, VIT\_18s0001g13930). We also detected 2-oxoglutarate and  $\text{Fe}^{2+}$ -dependent oxygenase superfamily proteins with members both up and downregulated (Supplementary Table S4 and Figure 6C), which catalyze the formation of hormones, such as ethylene, gibberellins, and pigments such as anthocyanidins and other flavones (Turnbull et al., 2004), again illustrating the broad spectrum and interconnectivity of responses observed.

## Modulation of Functions Involved in Cell Wall Remodeling

Since the thickening of secondary walls of xylem vessels is one of the most striking anatomical changes during PD onset (Figure 1C), we analyzed various host responses that could be classified as related to cell wall metabolism. We detected

several molecular systems working in cohort to produce the conspicuous thickening of xylem vessel walls. Some of the aforementioned enzymes of the phenylpropanoid biosynthesis pathways involved in the early steps of anthocyanin synthesis also generate precursors of monolignols. These are further processed and polymerized by oxidative enzymes, including peroxidases and copper-binding laccase-like polyphenol oxidases. Many laccases are highlighted in our data, with emphasis to VIT\_18s0001g00680 and VIT\_18s0122g00690 among 19 others, as well as six peroxidases including VIT\_14s0066g01670. Transcripts encoding 6 FAD-binding Berberine family proteins involved in monolignol oxidation were also accumulated, as well as for cytochrome P450-dependent monooxygenases known to be involved in lignin biosynthesis (Daniel et al., 2015), including VIT\_11s0078g00290 and VIT\_11s0065g00350 (cinnamate-4-hydroxylase) and also VIT\_04s0023g02900 (ferulic acid 5-hydroxylase).

Among other enzymatic functions involved with cell wall modification, we found reduced levels of two neighboring polygalacturonases (VIT\_01s0127g00850 and VIT\_01s0127g00870) and increased levels of a polygalacturonase inhibitor (VIT\_08s0007g07690) and of five dehiscence zone polygalacturonases orthologs of AT2G41850, eight pectin lyases including VIT\_07s0005g05520, 7 pectin invertase/methylesterase inhibitors such as VIT\_06s0009g02590, 11 cellulose synthases with emphasis to VIT\_02s0025g01940 (Figure 6D), the xyloglucan endotransglucosylase/ hydrolase VIT\_06s0061g00550, and two uclacyanins, all established in cell wall modification and/or lignin biosynthesis (Minic et al., 2009). Our proteomic analysis also identified upregulated gluconeogenic enzymes fructose-bisphosphate aldolase and alpha-glucan water dikinase, as well as TCA cycle enzymes



citrate synthase and isocitrate dehydrogenase, possibly contributing to generation of carbon backbones for cell wall thickening and ROS detoxification (Supplementary Table S8). Another protein with increased abundance was cinnamoyl-CoA reductase 1, known to participate in lignin biosynthesis (Ruel et al., 2009). Other modulated functions involved in cell wall modification functions encompassed eight expansins including VIT\_13s0067g02930, the exostosin VIT\_06s0061g00560, the beta galactosidase VIT\_11s0016g02200,

five beta glucosidases both positively and negatively regulated (VIT\_07s0005g00360 and VIT\_13s0064g01660, for example), two beta xylosidases (VIT\_05s0077g01280), three beta 1,3-glucanases including VIT\_06s0061g00100, and four hydroxyproline-rich LEA proteins (VIT\_04s0069g01010 in greater abundance). DE transcripts also include 18 UDP-glucosyl transferases with emphasis to VIT\_06s0004g07230, UDP-glucosyl epimerase VIT\_02s0025g04210, and the UDP-glucosyl dehydrogenase VIT\_17s0000g06960, all involved

in callose formation and deposition (Ellinger and Voigt, 2014).

## DISCUSSION

Much of our understanding of the plant responses to bacterial pathogens has been interpreted by the gene-for-gene response mediated by the secretion of type 3 effectors and their interaction with host R-proteins as a means to identify sources of resistance genes. Yet *X. fastidiosa* is a very successful pathogen for many economically important crops despite lacking a type 3 secretion system (T3SS) and other elicitors of plant immunity such as flagella. It uses, however, a T2SS that has been shown to secrete a number of hydrolytic enzymes that correlate to the observed disease symptoms. Our data further depict a wealth of molecular details of its complex interaction with grapevines leading to PD that suggest: (1) activation of a complex defense response that includes both pathogen- and damage-associated molecular pattern (PAMP/DAMP)-triggered immunity (PTI), but impairment of downstream salicylic acid-mediated immune response; (2) chronic oxidative stress despite activation of antioxidant metabolism; and (3) intensive cell wall remodeling and lignification, which can lead to reduced sap flow and increased water and nutrient deficiency. This is an expansion from the previous transcriptome investigation of the early events of PD development performed by Choi et al. (2013), in which the molecular events leading to water stress were also detected.

Given the xylem-dwelling characteristic of the pathogen, deep within the plant, and its ability to form biofilms and secrete virulence factors, pathogen clearance is not achieved in susceptible hosts despite the array of responses observed, including PR proteins and phytoalexins. Our data expand the set of PR proteins previously detected in the xylem sap of *X. fastidiosa*-infected grapevines (Chakraborty et al., 2016) further reinforcing their importance in the defense response. Though fastidious in growth, *X. fastidiosa* can actively migrate with and against xylem sap flow (Meng et al., 2005) and colonize the plant systemically, overcoming host strategies to halt pathogen proliferation. Callose and tylose formation are commonly seen throughout affected branches, and symptoms on leaves are not always associated with *X. fastidiosa* presence on the vicinity of scorched areas, suggesting the recognition and response to secreted effectors, PAMPs, and/or DAMPs (Gambetta et al., 2007; Nascimento et al., 2016; Rapicavoli et al., 2018). These signals might be associated to long range outer membrane vesicles, bacterial molecular signals, or products of host tissue degradation (Lindow et al., 2014; Nascimento et al., 2016; Rapicavoli et al., 2018). Pathogen and damage perception is suggested by induction of many LRR-RLKs, as well as downstream MAPKs and TFs of WRKY, and other types, all important for PTI (Banerjee and Roychoudhury, 2015). These receptors are the first layer of pathogen perception on the plant cell surface and directly or indirectly activate MAPK signaling cascades that affect a large group of TFs controlling the response circuitry (Zipfel, 2014). A large group of CDS encoding a wide array of ankyrin-repeat proteins were also strongly

modulated, which have been suggested to be intermediates between membrane-kinase receptors and downstream MAPKs (Yang et al., 2012), and were shown to increase resistance against bacterial blight caused by *Xanthomonas oryzae* (Wang et al., 2006) among other examples. It would be interesting to identify specific elicitors and signaling molecules recognized as PAMPs in *X. fastidiosa* pathosystems, and their cognate receptor pairs, as done, for example, between grapevine and *Burkholderia phytofirmans* (Trouvelot et al., 2014). In the case of *Xylella* that lacks flagella, this might help the pathogen to reduce the amount of immunogenic epitopes, as shown in other *Xanthomonads* that can regulate their flagellin biosynthesis, express multiple flagellin types, shed, or completely lack flagella (Darrasse et al., 2013), affecting their immunogenicity. We are currently pursuing this by expressing candidate *Xylella* proteins in grapevines, in order to better understand the host response to specific elicitors and assess whether PD symptoms derive from host responses to pathogen perception or from activity of the pathogen's effectors. Another interesting aspect of this plant-microbe crosstalk is the strong induction of GABA. Besides being part of the host response, it can also be used by bacterial quorum sensing, as exemplified by *Agrobacterium tumefaciens* in tobacco (Chevrot et al., 2006). Its precise effect on *X. fastidiosa* cellular behavior remains to be verified.

Our work also suggests that grapevines susceptible to PD go through intensive oxidative stress during disease development, as sources of ROS from PTI and also from photooxidative stress were evident in our data. ROS-generating systems were activated, such as germins, cupins, RBOH, and copper amine oxidases, possibly as a means to restrict pathogen proliferation. On the other hand, chronic exposure to high ROS levels is also detrimental to the host, as suggested by several ROS-scavenging strategies activated, among which we highlight the iron-sequestration ferritin nanocages, phenylpropanoid biosynthesis pathways, plus other proteins, and many metabolites aforementioned. Iron is needed to produce chlorophyll (Kumar and Soll, 2000), and hence its persistent chelation and consequent deficiency might intensify the conspicuous chlorosis symptoms and reduction of photosynthetic activity associated with PD. Our work provides further detail of this complex crosstalk between responses to pathogen and oxidative and drought stresses, as previously investigated in *A. thaliana* (Huang et al., 2008).

The intensive secondary metabolism modulation occurring during disease progression involves phytoalexin pathways (stilbene synthases) and also more macroscopic features such as cell wall thickening by phenylpropanoid and lignin biosynthesis. Although these are known components of grapevine's defense arsenal (Vannozzi et al., 2012), *X. fastidiosa* is able to evade these defense mechanisms and reach sufficient population thresholds that induce bacterial aggregation, biofilm formation, and efficient vector acquisition and transmission (Almeida et al., 2012). Exploring whether the upregulated metabolites can limit the bacterium's ability to reach such thresholds might be a promising way for delaying or preventing vectored disease transmission. Other insights toward rational disease control can come from the characterization of host susceptibility genes that can be exploited by the pathogen to harness nutrients. The nodulins identified in

this work are strong candidates in this path, as are the enzymes involved in salicylic acid metabolism that displayed decreased abundance possibly delaying immune responses. The sustained jasmonic acid-related responses coupled with inhibition of salicylic acid-related responses even after 12 weeks post-infection suggest that grapevines recognize *X. fastidiosa* as a necrotroph. Interestingly, this pattern of responses has also been observed in resistant citrus inoculated with *X. fastidiosa* during early stages of infection (1 day post-infection; Rodrigues et al., 2013). However, in the case of Thompson seedless grapevines, pathogen clearance is not attained and chronic exposure to oxidative stress combined with the virulence arsenal of the pathogen leads to plant death. Further investigations of particular pathways/analytes including more time-points will clarify the details and dynamics of the observed responses. As *Xylella* pathogens can switch between a more aggregated (biofilm) to a solitary/planktonic bioform, identifying the specific responses associated with each bioform is also warranted.

## CONCLUSION

Our work provides a set of molecular markers modulated by PD onset, as detected by different omics strategies. Our work attests the usefulness of using different omics strategies to dissect a pathosystem. Although some general features are clear on all of them such as the oxidative stress the host is going through, many are only evident in one or two of the techniques used. Integrating the omics data still presents many challenges, such as uniform annotation and nomenclature used by different tools and databases. The very different depth of the data, such as between the transcriptome and the proteome, also limit the extent to which an integrated analysis can be done. Nonetheless, we see a trend toward greater depth on all omics approaches as equipment and protocols advance. At this stage, our data already enable the selection among the various paralogs of given functions, for example, among the LRR-RLKs, thaumatin-like proteins, 2-oxoglutarate oxygenases, and cellulose synthases, as shown in **Figures 5, 6**. Understanding which paralogs respond most intensively to PD provides a valuable resource for targeted future investigations. The molecular markers for PD symptoms include proteins of various families plus antioxidant and antimicrobial metabolites that can now be further explored individually to evaluate their potential in disease detection and resistance. An integrated overview of the most intense responses detected is presented in **Figure 8**. Comparative studies with

the available breeding germplasm, analyzing specifically the markers highlighted herein, and engineering specific enhanced or reduced gene functions in order to increase resistance or reduce detrimental immune responses by silencing susceptibility genes are promising ways to address this.

## AVAILABILITY OF DATA

Raw transcriptome data are available for download at NCBI SRA BioProject # PRJNA390670. Filtered and normalized transcriptome data, along with proteome, metabolome data are available for download as supplementary material.

## AUTHOR CONTRIBUTIONS

RN, DC, LG, and AD designed the research. PZ, RN, HG, MP, and SC performed the research, data analysis, and interpretation. PZ and AD wrote the manuscript. All authors read and approved the final manuscript.

## FUNDING

This work was funded by grants PVE Capes 88881.064973/2014-01 and from California Department of Food and Agriculture Pierce's Disease Board. The funding bodies did not play any role in the design of the study and collection, analysis, and interpretation of the data and in writing the manuscript.

## ACKNOWLEDGMENTS

We thank Dr. Brett S. Phinney at UC Davis Proteomics Core Facility, Dr. David Dolan for his contributions to experimental procedures in the greenhouse, and Dr. Robson Souza and Dr. Joaquim Martins for help with high-throughput Blast analysis.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.00771/full#supplementary-material>

## REFERENCES

- Abou-Mansour, E., Debieux, J. L., Ramirez-Suero, M., Benard-Gellon, M., Magnin-Robert, M., Spagnolo, A., et al. (2015). Phytotoxic metabolites from *Neofusicoccum parvum*, a pathogen of *Botryosphaeria* dieback of grapevine. *Phytochemistry* 115, 207–215. doi: 10.1016/j.phytochem.2015.01.012
- Agüero, C. B., Uratsu, S. L., Greve, C., Powell, A. L., Labavitch, J. M., Meredith, C. P., et al. (2005). Evaluation of tolerance to Pierce's disease and *Botrytis* in transgenic plants of *Vitis vinifera* L. expressing the pear PGIP gene. *Mol. Plant Pathol.* 6, 43–51. doi: 10.1111/j.1364-3703.2004.00262.x
- Almeida, R. P., Killiny, N., Newman, K. L., Chatterjee, S., Ionescu, M., and Lindow, S. E. (2012). Contribution of *rpfB* to cell-to-cell signal synthesis, virulence, and vector transmission of *Xylella fastidiosa*. *Mol. Plant Microbe Interact.* 25, 453–462. doi: 10.1094/MPMI-03-11-0074
- Aoki, T., Aoki, Y., Ishiai, S., Otaguro, M., and Suzuki, S. (2016). Impact of *Bacillus cereus* NRKT on grape ripe rot disease through resveratrol synthesis in berry skin. *Pest Manag. Sci.* 73, 174–180. doi: 10.1002/ps.4283
- Ardales, E. Y., Moon, S. J., Park, S. R., Shin, D., Noh, T. H., and Byun, M. O. (2009). Inactivation of *argG*, encoding argininosuccinate synthetase from *Xanthomonas oryzae* pv. *oryzae*, is involved in bacterial growth and virulence in planta. *Can. J. Plant Pathol.* 31, 368–374. doi: 10.1080/07060660909507611



- Banerjee, A., and Roychoudhury, A. (2015). WRKY proteins: signaling and regulation of expression during abiotic stress responses. *ScientificWorldJournal* 2015:807560. doi: 10.1155/2015/807560
- Burger, J. T., and Maree, H. J. (2015). Metagenomic next-generation sequencing of viruses infecting grapevines. *Methods Mol. Biol.* 1302, 315–330. doi: 10.1007/978-1-4939-2620-6\_23
- Chakraborty, S., Nascimento, R., Zaini, P. A., Gouran, H., Rao, B. J., Goulart, L. R., et al. (2016). Sequence/structural analysis of xylem proteome emphasizes pathogenesis-related proteins, chitinases and beta-1, 3-glucanases as key players in grapevine defense against *Xylella fastidiosa*. *PeerJ* 4:e2007. doi: 10.7717/peerj.2007
- Chatterjee, S., Almeida, R. P., and Lindow, S. (2008). Living in two worlds: the plant and insect lifestyles of *Xylella fastidiosa*. *Annu. Rev. Phytopathol.* 46, 243–271. doi: 10.1146/annurev.phyto.45.062806.094342
- Chevrot, R., Rosen, R., Haudecoeur, E., Cirou, A., Shelp, B. J., Ron, E., et al. (2006). GABA controls the level of quorum-sensing signal in *Agrobacterium tumefaciens*. *Proc. Natl. Acad. Sci. U.S.A.* 103, 7460–7464. doi: 10.1073/pnas.0600313103
- Choi, M. S., Kim, W., Lee, C., and Oh, C. S. (2013). Harpins, multifunctional proteins secreted by gram-negative plant-pathogenic bacteria. *Mol. Plant Microbe Interact.* 26, 1115–1122. doi: 10.1094/MPMI-02-13-0050-CR
- Coleman, S. T., Fang, T. K., Rovinsky, S. A., Turano, F. J., and Moye-Rowley, W. S. (2001). Expression of a glutamate decarboxylase homologue is required for normal oxidative stress tolerance in *Saccharomyces cerevisiae*. *J. Biol. Chem.* 276, 244–250. doi: 10.1074/jbc.M007103200
- Couee, I., Sulmon, C., Gouesbet, G., and El Amrani, A. (2006). Involvement of soluble sugars in reactive oxygen species balance and responses to oxidative stress in plants. *J. Exp. Bot.* 57, 449–459. doi: 10.1093/jxb/erj027
- Czemmel, S., Galarneau, E. R., Travadon, R., McElrone, A. J., Cramer, G. R., and Baumgartner, K. (2015). Genes expressed in grapevine leaves reveal latent wood infection by the fungal pathogen *Neofusicoccum parvum*. *PLoS One* 10:e0121828. doi: 10.1371/journal.pone.0121828
- Dadakova, K., Havelkova, M., Kurkova, B., Tlojkova, I., Kasparovsky, T., Zdrahal, Z., et al. (2015). Proteome and transcript analysis of *Vitis vinifera* cell cultures subjected to *Botrytis cinerea* infection. *J. Proteomics* 119, 143–153. doi: 10.1016/j.jprot.2015.02.001
- Dandekar, A. M., Gouran, H., Ibanez, A. M., Uratsu, S. L., Aguero, C. B., McFarland, S., et al. (2012). An engineered innate immune defense protects grapevines from Pierce disease. *Proc. Natl. Acad. Sci. U.S.A.* 109, 3721–3725. doi: 10.1073/pnas.1116027109
- Daniel, B., Pavkov-Keller, T., Steiner, B., Dordic, A., Gutmann, A., Nidetzky, B., et al. (2015). Oxidation of monolignols by members of the berberine bridge enzyme family suggests a role in plant cell wall metabolism. *J. Biol. Chem.* 290, 18770–18781. doi: 10.1074/jbc.M115.659631
- Darrasse, A., Carrere, S., Barbe, V., Boureau, T., Arrieta-Ortiz, M. L., Bonneau, S., et al. (2013). Genome sequence of *Xanthomonas fuscans* subsp. *fuscans* strain 4834-R reveals that flagellar motility is not a general feature of xanthomonads. *BMC Genomics* 14:761. doi: 10.1186/1471-2164-14-761
- Ellinger, D., and Voigt, C. A. (2014). Callose biosynthesis in *Arabidopsis* with a focus on pathogen response: what we have learned within the last decade. *Ann. Bot.* 114, 1349–1358. doi: 10.1093/aob/mcu120
- Falshaw, A., Hart, J. B., and Tyler, P. C. (2000). New syntheses of 1D- and 1L-1,2-anhydro-myo-inositol and assessment of their glycosidase inhibitory activities. *Carbohydr. Res.* 329, 301–308. doi: 10.1016/S0008-6215(00)00192-0
- Fenyo, D., and Beavis, R. C. (2003). A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes. *Anal. Chem.* 75, 768–774. doi: 10.1021/ac0258709
- Fiehn, O., Wohlgemuth, G., and Scholz, M. (2005). "Setup and annotation of metabolomic experiments by integrating biological and mass spectrometric metadata," in *Data Integration in the Life Sciences. DILS 2005 Lecture Notes in Bioinformatics*, Vol. 3615 eds B. Ludäscher, and L. Raschid (Berlin: Springer), 224–239. doi: 10.1007/11530084\_18
- Gambetta, G. A., Fei, J., Rost, T. L., and Matthews, M. A. (2007). Leaf scorch symptoms are not correlated with bacterial populations during Pierce's disease. *J. Exp. Bot.* 58, 4037–4046. doi: 10.1093/jxb/erm260
- Ghezylbash, G. R., Nahvi, I., and Rabbani, M. (2012). Comparative inhibitory effect of xylitol and erythritol on the growth and biofilm formation of oral *Streptococci*. *Afr. J. Microbiol. Res.* 6, 4404–4408. doi: 10.5897/AJMR11.1122
- Gouran, H., Gillespie, H., Nascimento, R., Chakraborty, S., Zaini, P. A., Jacobson, A., et al. (2016). The secreted protease PrtA controls cell growth, biofilm formation and pathogenicity in *Xylella fastidiosa*. *Sci. Rep.* 6:31098. doi: 10.1038/srep31098
- Hasanuzzaman, M., Nahar, K., Hossain, M. S., Mahmud, J. A., Rahman, A., Inafuku, M., et al. (2017). Coordinated actions of glyoxalase and antioxidant defense systems in conferring abiotic stress tolerance in plants. *Int. J. Mol. Sci.* 18:200. doi: 10.3390/ijms18010200
- Huang, D., Wu, W., Abrams, S. R., and Cutler, A. J. (2008). The relationship of drought-related gene expression in *Arabidopsis thaliana* to hormonal and environmental factors. *J. Exp. Bot.* 59, 2991–3007. doi: 10.1093/jxb/ern155
- Jaillon, O., Aury, J. M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., et al. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449, 463–467. doi: 10.1038/nature06148
- Kadota, Y., Shirasu, K., and Zipfel, C. (2015). Regulation of the NADPH Oxidase RBOHD During Plant Immunity. *Plant Cell Physiol.* 56, 1472–1480. doi: 10.1093/pcp/pcv063
- Katam, R., Chibanguza, K., Latinwo, L. M., and Smith, D. (2015). Proteome biomarkers in Xylem Reveal Pierce's Disease Tolerance in Grape. *J. Proteomics Bioinform.* 8, 217–224. doi: 10.4172/jpb.1000372
- Kelloniemi, J., Trouvelot, S., Heloir, M. C., Simon, A., Dalmis, B., Frettinger, P., et al. (2015). Analysis of the molecular dialogue between gray mold (*Botrytis cinerea*) and grapevine (*Vitis vinifera*) reveals a clear shift in defense mechanisms during berry ripening. *Mol. Plant Microbe Interact.* 28, 1167–1180. doi: 10.1094/MPMI-02-15-0039-R
- Kind, T., Wohlgemuth, G., Lee, D. Y., Lu, Y., Palazoglu, M., Shahbaz, S., et al. (2009). FiehnLib: mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Anal. Chem.* 81, 10038–10048. doi: 10.1021/ac9019522
- Kinnersley, A. M., and Turano, F. J. (2000). Gamma aminobutyric acid (GABA) and plant responses to stress. *Crit. Rev. Plant Sci.* 19, 479–509. doi: 10.1080/07352680091139277
- 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
- Kumar, A. M., and Soll, D. (2000). Antisense HEMA1 RNA expression inhibits heme and chlorophyll biosynthesis in *Arabidopsis*. *Plant Physiol.* 122, 49–56. doi: 10.1104/pp.122.1.49
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Lin, H., Doddapaneni, H., Takahashi, Y., and Walker, M. A. (2007). Comparative analysis of ESTs involved in grape responses to *Xylella fastidiosa* infection. *BMC Plant Biol.* 7:8. doi: 10.1186/1471-2229-7-8
- Lindow, S., Newman, K., Chatterjee, S., Baccari, C., Lavarone, A. T., and Ionescu, M. (2014). Production of *Xylella fastidiosa* diffusible signal factor in transgenic grape causes pathogen confusion and reduction in severity of Pierce's disease. *Mol. Plant Microbe Interact.* 27, 244–254. doi: 10.1094/MPMI-07-13-0197-FI
- Martelli, G. P., Boscia, D., Porcelli, F., and Saponari, M. (2016). The olive quick decline syndrome in south-east Italy: a threatening phytosanitary emergency. *Eur. J. Plant Pathol.* 144, 235–243. doi: 10.1007/s10658-015-0784-7
- Mehterov, N., Balazadeh, S., Hille, J., Toneva, V., Mueller-Roeber, B., and Gechev, T. (2012). Oxidative stress provokes distinct transcriptional responses in the stress-tolerant *atr7* and stress-sensitive *loh2* *Arabidopsis thaliana* mutants as revealed by multi-parallel quantitative real-time PCR analysis of ROS marker and antioxidant genes. *Plant Physiol. Biochem.* 59, 20–29. doi: 10.1016/j.plaphy.2012.05.024
- Meng, Y., Li, Y., Galvani, C. D., Hao, G., Turner, J. N., Burr, T. J., et al. (2005). Upstream migration of *Xylella fastidiosa* via pilus-driven twitching motility. *J. Bacteriol.* 187, 5560–5567. doi: 10.1128/JB.187.16.5560-5567.2005
- Mi, H., Poudel, S., Muruganujan, A., Casagrande, J. T., and Thomas, P. D. (2016). PANTHER version 10: expanded protein families and functions, and analysis tools. *Nucleic Acids Res.* 44, D336–D342. doi: 10.1093/nar/gkv1194

- Minic, Z., Jamet, E., San-Clemente, H., Pelletier, S., Renou, J. P., Rihouey, C., et al. (2009). Transcriptomic analysis of Arabidopsis developing stems: a close-up on cell wall genes. *BMC Plant Biol.* 9:6. doi: 10.1186/1471-2229-9-6
- Mittler, R. (2017). ROS are good. *Trends Plant Sci.* 22, 11–19. doi: 10.1016/j.tplants.2016.08.002
- Mizuno, D., Higuchi, K., Sakamoto, T., Nakanishi, H., Mori, S., and Nishizawa, N. K. (2003). Three nicotianamine synthase genes isolated from maize are differentially regulated by iron nutritional status. *Plant Physiol.* 132, 1989–1997. doi: 10.1104/pp.102.019869
- Nascimento, R., Gouran, H., Chakraborty, S., Gillespie, H. W., Almeida-Souza, H. O., Tu, A., et al. (2016). The Type II secreted Lipase/Esterase LesA is a key virulence factor required for *Xylella fastidiosa* pathogenesis in grapevines. *Sci. Rep.* 6:18598. doi: 10.1038/srep18598
- Norholm, M. H., Nour-Eldin, H. H., Brodersen, P., Mundy, J., and Halkier, B. A. (2006). Expression of the Arabidopsis high-affinity hexose transporter STP13 correlates with programmed cell death. *FEBS Lett.* 580, 2381–2387. doi: 10.1016/j.febslet.2006.03.064
- Rapicavoli, J. N., Blanco-Ulate, B., Muszynski, A., Figueroa-Balderas, R., Morales-Cruz, A., Azadi, P., et al. (2018). Lipopolysaccharide O-antigen delays plant innate immune recognition of *Xylella fastidiosa*. *Nat. Commun.* 9:390. doi: 10.1038/s41467-018-02861-5
- Roberts, A., Pimentel, H., Trapnell, C., and Pachter, L. (2011). Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* 27, 2325–2329. doi: 10.1093/Bioinformatics/Btr355
- 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
- Rodrigues, C. M., de Souza, A. A., Takita, M. A., Kishi, L. T., and Machado, M. A. (2013). RNA-Seq analysis of *Citrus reticulata* in the early stages of *Xylella fastidiosa* infection reveals auxin-related genes as a defense response. *BMC Genomics* 14:676. doi: 10.1186/1471-2164-14-676
- Roper, M. C., Greve, L. C., Warren, J. G., Labavitch, J. M., and Kirkpatrick, B. C. (2007). *Xylella fastidiosa* requires polygalacturonase for colonization and pathogenicity in *Vitis vinifera* grapevines. *Mol. Plant Microbe Interact.* 20, 411–419. doi: 10.1094/MPMI-20-4-0411
- Ruel, K., Berrio-Sierra, J., Derikvand, M. M., Pollet, B., Thevenin, J., Lapierre, C., et al. (2009). Impact of CCR1 silencing on the assembly of lignified secondary walls in *Arabidopsis thaliana*. *New Phytol.* 184, 99–113. doi: 10.1111/j.1469-8137.2009.02951.x
- Sana, T. R., Fischer, S., Wohlgemuth, G., Katrekar, A., Jung, K. H., Ronald, P. C., et al. (2010). Metabolomic and transcriptomic analysis of the rice response to the bacterial blight pathogen *Xanthomonas oryzae* pv. *oryzae*. *Metabolomics* 6, 451–465. doi: 10.1007/s11306-010-0218-7
- Savoi, S., Wong, D. C., Arapitsas, P., Miculan, M., Bucchetti, B., Peterlunger, E., et al. (2016). Transcriptome and metabolite profiling reveals that prolonged drought modulates the phenylpropanoid and terpenoid pathway in white grapes (*Vitis vinifera* L.). *BMC Plant Biol.* 16:67. doi: 10.1186/s12870-016-0760-1
- Schaller, F., Biesgen, C., Mussig, C., Altmann, T., and Weiler, E. W. (2000). 12-Oxophytodienoate reductase 3 (OPR3) is the isoenzyme involved in jasmonate biosynthesis. *Planta* 210, 979–984. doi: 10.1007/s004250050706
- Scholz, S. S., Malabarba, J., Reichelt, M., Heyer, M., Ludewig, F., and Mithofer, A. (2017). Evidence for GABA-induced systemic GABA accumulation in *Arabidopsis* upon wounding. *Front. Plant Sci.* 8:388. doi: 10.3389/fpls.2017.00388
- Sels, J., Mathys, J., De Coninck, B. M., Cammue, B. P., and De Bolle, M. F. (2008). Plant pathogenesis-related (PR) proteins: a focus on PR peptides. *Plant Physiol. Biochem.* 46, 941–950. doi: 10.1016/j.plaphy.2008.06.011
- Sengupta, S., Mukherjee, S., Basak, P., and Majumder, A. L. (2015). Significance of galactinol and raffinose family oligosaccharide synthesis in plants. *Front. Plant Sci.* 6:656. doi: 10.3389/fpls.2015.00656
- Sun, Q., Sun, Y., Walker, M. A., and Labavitch, J. M. (2013). Vascular occlusions in grapevines with Pierce's disease make disease symptom development worse. *Plant Physiol.* 161, 1529–1541. doi: 10.1104/pp.112.208157
- Tello-Ruiz, M. K., Stein, J., Wei, S., Preece, J., Olson, A., Naithani, S., et al. (2016). Gramene 2016: comparative plant genomics and pathway resources. *Nucleic Acids Res.* 44, D1133–D1140. doi: 10.1093/nar/gkv1179
- Thimm, O., Blasing, O., Gibon, Y., Nagel, A., Meyer, S., Kruger, P., et al. (2004). MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.* 37, 914–939. doi: 10.1111/j.1365-3113.2004.02016.x
- Trouvelot, S., Heloir, M. C., Poinssot, B., Gauthier, A., Paris, F., Guiller, C., et al. (2014). Carbohydrates in plant immunity and plant protection: roles and potential application as foliar sprays. *Front. Plant Sci.* 5:592. doi: 10.3389/fpls.2014.00592
- Turnbull, J. J., Nakajima, J., Welford, R. W., Yamazaki, M., Saito, K., and Schofield, C. J. (2004). Mechanistic studies on three 2-oxoglutarate-dependent oxygenases of flavonoid biosynthesis: anthocyanidin synthase, flavonol synthase, and flavanone 3beta-hydroxylase. *J. Biol. Chem.* 279, 1206–1216. doi: 10.1074/jbc.M309228200
- Vannozzi, A., Dry, I. B., Fasoli, M., Zenoni, S., and Lucchin, M. (2012). Genome-wide analysis of the grapevine stilbene synthase multigenic family: genomic organization and expression profiles upon biotic and abiotic stresses. *BMC Plant Biol.* 12:130. doi: 10.1186/1471-2229-12-130
- Wallis, C. M., Wallingford, A. K., and Chen, J. (2013). Grapevine rootstock effects on scion sap phenolic levels, resistance to *Xylella fastidiosa* infection, and progression of Pierce's disease. *Front. Plant Sci.* 4:502. doi: 10.3389/fpls.2013.00502
- Wang, J., Feng, J., Jia, W., Chang, S., Li, S., and Li, Y. (2015). Lignin engineering through laccase modification: a promising field for energy plant improvement. *Biotechnol. Biofuels* 8:145. doi: 10.1186/s13068-015-0331-y
- Wang, Y. S., Pi, L. Y., Chen, X., Chakraborty, P. K., Jiang, J., De Leon, A. L., et al. (2006). Rice XA21 binding protein 3 is a ubiquitin ligase required for full Xa21-mediated disease resistance. *Plant Cell* 18, 3635–3646. doi: 10.1105/tpc.106.046730
- Xie, D. Y., Sharma, S. B., Paiva, N. L., Ferreira, D., and Dixon, R. A. (2003). Role of anthocyanidin reductase, encoded by BANYULS in plant flavonoid biosynthesis. *Science* 299, 396–399. doi: 10.1126/science.1078540
- Yang, Y., Zhang, Y., Ding, P., Johnson, K., and Li, X. (2012). The ankyrin-repeat transmembrane protein BDA1 functions downstream of the receptor-like protein SNC2 to regulate plant immunity. *Plant Physiol.* 159, 1857–1865. doi: 10.1104/pp.112.197152
- Zipfel, C. (2014). Plant pattern-recognition receptors. *Trends Immunol.* 35, 345–351. doi: 10.1016/j.it.2014.05.004

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

Copyright © 2018 Zaini, Nascimento, Gouran, Cantu, Chakraborty, Phu, Goulart and Dandekar. 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.



# Metabolome Integrated Analysis of High-Temperature Response in *Pinus radiata*

Mónica Escandón<sup>1\*†</sup>, Mónica Meijón<sup>1,2</sup>, Luis Valledor<sup>1,2</sup>, Jesús Pascual<sup>3</sup>, Gloria Pinto<sup>4</sup> and María Jesús Cañal<sup>1,2\*</sup>

<sup>1</sup> Plant Physiology, Department of Organisms and Systems Biology, Faculty of Biology, University of Oviedo, Oviedo, Spain,

<sup>2</sup> Plant Biotechnology Unit, University Institute of Biotechnology of Asturias (IUBA), Oviedo, Spain, <sup>3</sup> Molecular Plant Biology, Department of Biochemistry, University of Turku, Turku, Finland, <sup>4</sup> Department of Biology and CESAM, University of Aveiro, Aveiro, Portugal

## OPEN ACCESS

### Edited by:

Atsushi Fukushima,  
RIKEN, Japan

### Reviewed by:

Manoj Kumar,  
University of Technology Sydney,  
Australia  
Saleh Alseekh,  
Max Planck Institute of Molecular  
Plant Physiology (MPG), Germany

### \*Correspondence:

Mónica Escandón  
escandonmonica@uniovi.es;  
m.escandon.martinez@gmail.com  
María Jesús Cañal  
mjcanal@uniovi.es

### † Present address:

Mónica Escandón,  
Department of Biology and CESAM,  
University of Aveiro, Aveiro, Portugal

### Specialty section:

This article was submitted to  
Plant Systems and Synthetic Biology,  
a section of the journal  
Frontiers in Plant Science

Received: 07 November 2017

Accepted: 29 March 2018

Published: 17 April 2018

### Citation:

Escandón M, Meijón M, Valledor L,  
Pascual J, Pinto G and Cañal MJ  
(2018) Metabolome Integrated  
Analysis of High-Temperature  
Response in *Pinus radiata*.  
Front. Plant Sci. 9:485.  
doi: 10.3389/fpls.2018.00485

The integrative omics approach is crucial to identify the molecular mechanisms underlying high-temperature response in non-model species. Based on future scenarios of heat increase, *Pinus radiata* plants were exposed to a temperature of 40°C for a period of 5 days, including recovered plants (30 days after last exposure to 40°C) in the analysis. The analysis of the metabolome using complementary mass spectrometry techniques (GC-MS and LC-Orbitrap-MS) allowed the reliable quantification of 2,287 metabolites. The analysis of identified metabolites and highlighter metabolic pathways across heat time exposure reveal the dynamism of the metabolome in relation to high-temperature response in *P. radiata*, identifying the existence of a turning point (on day 3) at which *P. radiata* plants changed from an initial stress response program (shorter-term response) to an acclimation one (longer-term response). Furthermore, the integration of metabolome and physiological measurements, which cover from the photosynthetic state to hormonal profile, suggests a complex metabolic pathway interaction network related to heat-stress response. Cytokinins (CKs), fatty acid metabolism and flavonoid and terpenoid biosynthesis were revealed as the most important pathways involved in heat-stress response in *P. radiata*, with zeatin riboside (ZR) and isopentenyl adenosine (iPA) as the key hormones coordinating these multiple and complex interactions. On the other hand, the integrative approach allowed elucidation of crucial metabolic mechanisms involved in heat response in *P. radiata*, as well as the identification of thermotolerance metabolic biomarkers (L-phenylalanine, hexadecanoic acid, and dihydromyricetin), crucial metabolites which can reschedule the metabolic strategy to adapt to high temperature.

**Keywords:** pine, heat-acclimation, metabolomics, multivariate integrative analyses, biomarkers

## INTRODUCTION

As a consequence of climate change, the intensity and frequency of extreme weather events, such as heat waves, are projected to increase, these being one of the major global risks (EEA, 2012). Heat waves (when the temperature rises at least 5°C above normal values) can have a short duration (about a few days). Although plants are known to be able to respond to altered climatic conditions

(e.g., Schworer et al., 2014), in woody plants, and specifically the response and heat adaption rate have been barely studied. In forest ecosystems, climate change could have significant implications in timber production (Martinich et al., 2017). Therefore, in order to reach the future demand for wood products, it is necessary to focus research in improving the production, health, and performance of the commercially valued forest species, such as *Pinus radiata*, in future scenarios of increased temperature.

In plants, it has been reported that high temperature has negative effects in various physiological processes such as photosynthesis, primary and secondary metabolism, water relations, or lipid metabolism (Xu et al., 2006). Specifically, it is known that heat stress generates damage to the cell membrane, overproduction of reactive oxygen species (ROS), senescence, inhibition of photosynthesis, and cell death. However, plants have designed specific protection mechanisms to minimize and repair the damage caused by high temperatures in order to cope with heat waves that they will inevitably face during their lifespan.

When a plant perceives environmental stress, multiple signaling cascades are activated. These include the intricate cross-talk between the different plant hormones and other signaling pathways involving kinases and phosphatases, calcium, ROS, and lipids. Previous studies in *P. radiata* showed salicylic acid (SA) and abscisic acid (ABA) as crucial factors in the initial response to heat stress (Escandón et al., 2016), probably due to the rush of the plant to quickly regulate stomatal closure as observed in other species (Acharya and Assmann, 2009). Other phytohormones such as cytokinins (CKs) and indolacetic acid (IAA) seem to be more important for the acclimation and recovery of the plant (Skalák et al., 2016). Nevertheless, how these hormones interact and trigger changes in the plant metabolome to deal with heat stress is still unknown.

Environmental stress results in a reorganization of the metabolism in order to assure homeostasis, which is often accomplished by maintaining essential metabolism and synthesizing metabolites with stress-protective and signaling properties (Fernández de Simón et al., 2017). Changes in secondary metabolism are usually triggered by environmental stresses (Routaboul et al., 2012; Fernández de Simón et al., 2017) such as extreme temperatures, salinity, water availability, or high light intensity, being involved in most plant adaptive responses in combination with signaling mechanisms regulated by plant hormones. Phenolic compounds are one of the most important classes of secondary metabolites in plants as they play important roles in the response to high-temperature stress. Specifically, high-temperature stress promotes the production of phenolic compounds such as flavonoids, phenylpropanoids, anthocyanins, and lignins which are related to the suppression of stress-induced oxidation of most cell molecules (Wahid et al., 2007).

Lipid metabolism is also altered by high-temperature stress as a basic mechanism to control membrane fluidity, cell signaling, and movement of substances (Falcone et al., 2004). According to Grover et al. (2000) and Umesha (2005), the accumulation of highly saturated fatty acids might confer tolerance to high-temperature stress by means of reducing structural membrane fluidity when this is increased by environmental warmth.

Metabolomics profile analyses can be considered an important diagnostic tool for verifying the physiological responses of plant species to environmental changes and to understand the mechanisms behind the complex biological response (Gibbons et al., 2015; Zhang et al., 2015; Meijón et al., 2016), providing a snapshot of the physiological status of the plant in response to an environmental stress (Gibbons et al., 2015; Patel and Ahmed, 2015; Zhang et al., 2015). The feasibility of metabolome analysis for biomarker discovery relies on the assumption that metabolites are important players in biological systems (Monteiro et al., 2013) and stress situations cause drastic changes of metabolomics pathways, which are not new concepts (Shulaev et al., 2008; de Leonardis et al., 2015; Pascual et al., 2017). Currently, using MS-based platforms and combining different analytical technologies, it is possible to increase metabolome coverage. The use of GC-MS technique allows measuring the most of the primary metabolites, while LC-MS provide a better coverage of large hydrophobic metabolites predominant in secondary metabolisms (Doerfler et al., 2013). Both techniques together can assure a better metabolome, allowing the elicitation of a complete view of metabolic dynamics involved in the heat response of *P. radiata*. However, these kinds of studies also require a system biological approach using bioinformatics tools to understand their implications for cell function and to attach the missing connections between molecules and plant physiology (Bruggeman and Westerhoff, 2007; Meijón et al., 2016).

In this study, *P. radiata* plants were exposed to high temperatures aiming to mimic future scenarios of increased warmth. Metabolome and physiological data were analyzed comprehensively using a multivariable approach combining both sets of data. The analysis revealed the dynamic behavior of the metabolic and signaling transduction pathways, as well as connections among the pathways. Classical physiological measurements combined with cutting-edge technologies such as mass spectrometry-based analytical procedures for characterizing the variations in the metabolome revealed key metabolites related to high-temperature response in *P. radiata*. These key metabolites have a possible use as biomarkers in *P. radiata* and other species, due to the similarity of metabolites and basic metabolic pathways between very different species, while proteins, genes, and mRNAs are diversified from one species to another (Peng et al., 2015). Altogether, this work provides a deep knowledge of the response and acclimation process to heat stress in *P. radiata*, as well as the selection of possible universal thermotolerance biomarkers or crucial metabolites which can be further considered by breeders and forest managers.

## MATERIALS AND METHODS

### Plant Material and Experimental Design

The assay was conducted in a climate chamber under controlled conditions (Fitoclima 1200, Aralab). One-year-old *P. radiata* seedlings (plant size about  $33 \pm 4$  cm) in 1-dm<sup>3</sup> pots (blond peat:vermiculite, 1:1) were kept under a photoperiod of 16 h



(400  $\mu\text{mol m}^{-2}\cdot\text{s}^{-1}$ ) at 25°C and 50% relative humidity (RH), and 15°C and 60 % RH during the night period. The plants had been previously acclimated over a 1-month period inside the climate chamber, being watered with nutritive solution (NPK, 5:8:10).

Control plants (C) were collected before starting the heat exposure and were maintained at 25°C for the duration of the trial. Heat exposure treatment began with an increasing temperature gradient from 15 to 40°C over 5 h and maintained for 6 h. This experimental procedure was repeated for 5 days. Sampling was performed at: 3 h after 40°C was reached on day 1 (T1/2) and at the end of the 6-h heat exposure on day 1 (T1), day 2 (T2), day 3 (T3), and day 5 (T5). Plants were watered every day to 80% FC (full capacity) and weekly fertilized with a nutritive solution (NPK, 5:8:10). Plants of each treatment were allowed to recover for 1 month under the control conditions. Recovered plants (R) represent an intermediate exposure time (T3 recovered plants) because there was no significant difference (at the morpho-physiological level) between recovered plants exposed to the different heat exposures. This experimental design aimed to cover the entire stress sensing–response–adaptation process, increasing the density of analysis at short-term (T1/2, T1, T2), longer exposures (T5), and recovered plants (R), complementing previous short-term response (C, T1, and T3) analysis (Escandón et al., 2017).

Mature needles from each plant (16 plants/exposure) were sampled, cleaned with a moistened cloth, and immediately frozen in liquid nitrogen until metabolites were extracted. Pools of 3 plants for each biological replicate were performed. Data of physiological measurements – electrolyte leakage (EL), relative water content (RWC), maximum quantum efficiency of PSII (Fv/Fm), quantum yield of photosystem II photochemistry ( $\phi\text{PSII}$ ), malondialdehyde content (MDA), proline content, starch content, total soluble sugars (TSS), chlorophyll a content (Chla), chlorophyll b content (Chlb), and carotenoid content (Carot) – and fitohormone data – SA, indol-3-acetic acid (IAA), ABA, zeatin riboside (ZR), dihydrozeatin riboside (DHZR), gibberellin 7 (GA<sub>7</sub>), jasmonic acid (JA), gibberellin 9 (GA<sub>9</sub>), isopentenyl adenosine (iPA), isopentenyl adenine (iP), and castasterone (BK) – were taken from Escandón et al. (2016) for multivariable and integrative analysis.

## Metabolite Extraction

Metabolite extraction was performed according to Valledor et al. (2014a) using 100 mg of needle fresh weight. Briefly, samples (C, T1/2, T1, T2, T3, T5, R) were ground in liquid nitrogen and 600  $\mu\text{L}$  of cold (4°C) metabolite extraction solution – Methanol:Chloroform:H<sub>2</sub>O (2.5:1:0.5) – was immediately added to each tube. Then, samples were centrifuged at 20,000 g for 4 min at 4°C and the supernatant transferred to new tubes. Finally, 800  $\mu\text{L}$  of Chloroform:water (1:1) were added and the tubes vortexed and again centrifuged at 20,000 g for 4 min at 4°C. Two layers formed: an upper aqueous layer, containing the polar metabolites and a lower organic layer, containing the non-polar. Both fractions were transferred to new tubes and dried in a speed vac.

## Polar Metabolite Identification and Quantitation Using LC-Orbitrap-MS Analysis

The polar fraction of each sample was analyzed twice on an LC-Orbitrap-MS, first in positive ion mode and then in negative. A Dionex Ultimate 3000 (Thermo Fisher Scientific, United States) UHPLC was used and a LC-Orbitrap LTQ XL-MS system (controlled by Xcalibur version 2.2, Thermo Fisher Corporation) was run according to the procedure described in Escandón et al. (2017). The resolution and sensitivity of the Orbitrap were controlled by the injection of a mixed standard after the analysis of each batch, and the resolution was also checked with the aid of lock masses (phthalates). Blanks were also analyzed during the sequence.

LC-Orbitrap-MS raw data were processed and compared using MZmine software version 2.10 (Pluskal et al., 2010). MS1 spectra were filtered establishing a noise threshold at  $5.5\text{E}^{03}$  and minimum peak height at  $6\text{E}^{03}$  with a minimum time peak of 0.15 min. Peaks were smoothed and deconvoluted by using a local minimum search algorithm (98% chromatographic threshold, minimum retention range 5 min, minimum relative height of 90%, and minimum ratio top/edge of 1.2). Chromatograms were aligned using the RANSAC algorithm with a tolerance of 5 ppm of and 1.0 min retention time. Normalized peak areas were used for quantification, and their values were log transformed before statistical analyses (Supplementary Table S1a).

The individual peaks were identified following different approaches; the first step was performed against an *in-house* library (> 100 compounds) and manual annotation considering m/z and retention times. In the second step, masses were assigned using the KEGG, PubChem, METLIN, MassBank, HMDB, and Plantcyc databases as reported by Escandón et al. (2017) with built-in MZmine plugins with a 5 ppm threshold and considering as “identified” beyond doubt those metabolites that were defined after the comparison to our standard compound library or by a matching of MS/MS to the small number of plant compounds for which their MS/MS is available in public databases (Meijón et al., 2016); and as “tentatively assigned” those with molecular ions with exact masses corresponding to identified metabolites in databases. Metabolite identification against our library was confirmed by retention time (RT), mass, isotopic pattern, and ring double bond parameters. Supplementary Data S1 includes the detailed interpretation of experimental MS/MS spectra which support our tentative identifications of the candidate metabolites that were not identified beyond doubt in the first term.

## Non-polar Metabolite Identification and Quantitation Analysis

Non-polar metabolites were derivatized with 295  $\mu\text{L}$  tert-methyl-Butyl-Ether (MTBE) and 5  $\mu\text{L}$  of trimethylsulfonium hydroxide (TMSH) for 30 min at room temperature. The tubes were centrifuged for 3 min at 20,000 g to remove insoluble particles before transferring the supernatants to GC-microvials. GC-MS measurements were carried out following a previously developed procedure (Valledor et al., 2014b) on a triple quad instrument (TSQ Quantum GC; Thermo, United States). The

mass spectrometer was operated in electron-impact (EI) mode at 70 eV in a scan range of  $m/z$  40–600. Metabolites were identified based on their mass spectral characteristics and GC retention times through comparison with the retention times of reference compounds in an in-house reference library and the current version of Golm Metabolome Database (Hummel et al., 2007) using LC-Quant software (Supplementary Table S1b).

## Quantitative Real-Time PCR of Selected Genes

RNA was extracted from 100 mg of needle fresh weight as described by Valledor et al. (2014a). cDNA was obtained from 1,000 ng of RNA using the RevertAid kit (Thermo Scientific, United States) and random hexamers as primers following the manufacturer's instructions. Later, qPCR reactions were performed in a CFX Connect Real Time PCR machine (Bio-Rad) with SsoAdvanced Universal SYBR Green Supermix (Bio-Rad, United States); three biological and two analytical replicates were performed for each treatment.

*ACTIN* (*ACT*), *RIBOSOMAL PROTEIN 18S*, *GLYCERALDEHYDE 3-212 PHOSPHATE DEHYDROGENASE* (*GAPDH*), and *UBIQUITINE* (*UBI*) genes were tested as endogenous control employing geNorm following the criteria of Hellemans et al. (2007). *ACT* and *UBI* were the most stable and consequently selected as endogenous genes. Normalized Relative Quantities (NRQ) and Standard Errors of RQ were determined according to Hellemans et al. (2007). Primers were designed using transcript sequences available in a *P. radiata* in-house database obtained from RNA-Seq data (unpublished results). Detailed information about the primers used for qPCR experiments is available in Supplementary Table S2.

## Statistical and Bioinformatics Analysis

Five biological replicates were used for metabolites and physiological parameter statistical analysis. The procedures were conducted with the R programming language running under the open-source computer software R v2.15.2 (R Development Core Team, 2015) and RStudio (RStudio Team, 2015). Metabolome datasets were pre-processed following the recommendations of Valledor et al. (2014c). In brief, missed values were imputed using a k-nearest neighbors approach, and variables were filtered out if they were not present in all replicates of one treatment or in at least 45% of the analyzed samples. Data was transformed following a sample-centric approach followed by log transformation. Centered and scaled values (z-scores) were subjected to multivariate analysis and Heatmap clustering. The calculation of the number of common metabolites for all combinations of treatments and unique metabolites in a single treatment was performed using core functions of R.

Metabolomics pathways of each metabolite (Supplementary Table S3a) were searched against KEGG pathway maps (KEGG Mapper, Kanehisa et al., 2012) and *p*-values of each metabolomics pathways (Supplementary Table S3b) in MBROLE 2.0 (López-Ibáñez et al., 2016). Heat mapping was carried out using the Manhattan distance method to group metabolites in different

KEGG pathways with an MBROLE FDR correction of less than 0.05. Multivariate analysis of metabolites and physiological parameters (Supplementary Table S4) were conducted with mixOmics (Lê Cao et al., 2009) using Principal Component Analysis (PCA), Sparse Partial Least Squares (sPLS), and network analyses. The normalization of the datasets was performed before combining them. sPLS algorithm was used to find correlations between predictor (metabolites matrix) and response variables (physiological parameters) and its graphic representation in the network analysis. Network topology was defined after applying sPLS regression using the function `network()` of the mixOmics package and filtered (only edges equal or higher than  $|0.60|$  were maintained) in Cytoscape v.3.3.0 (Cline et al., 2007). Univariate analyses were conducted: one-way ANOVA,  $p < 0.05$  for metabolites and physiological parameters and Student's *t*-test,  $p < 0.1$  for qPCR analysis. Graphics were plotted employing ggplot2 (Wickham, 2009) and pheatmap (Kolde, 2015).

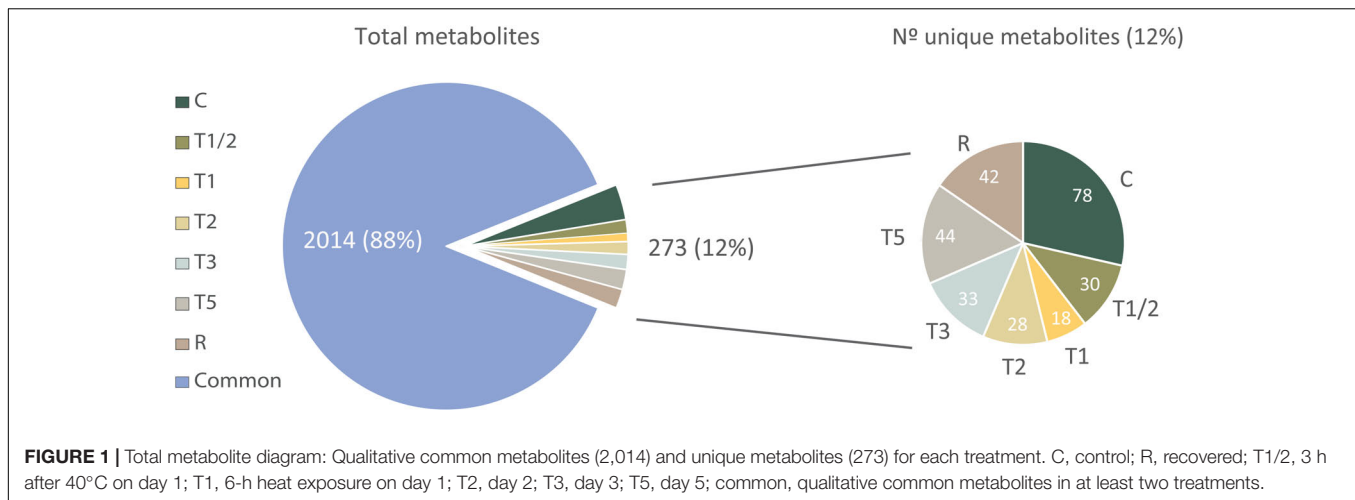
## RESULTS

### Characterization of the Metabolome During Heat Treatment

The combination of LC-Orbitrap-MS for polar and GC-MS for non-polar metabolites allowed the reliable quantification of a total of 2,287 ions based on the obtained  $m/z$  and retention times (Supplementary Table S1). These different ions can be considered different metabolites since each combination of  $m/z$  and retention time should be unique for each metabolite and spatial conformation. The fusion of a customized searching/identification algorithm based on in-house and public databases resulted in the unequivocal identification of 41 ions (identical matches to our compound library) and 747 ions that were tentatively assigned after comparing its very accurate mass against reference compound databases (Supplementary Table S1). The combination of both ionization modes (GC-MS and LC-Orbitrap-MS) gave a broad characterization of the pine metabolomes during heat-induced response, which covered most of the primary and secondary metabolism pathways.

Metabolome analysis showed that most of the metabolites are present in all treatments or at least in two different treatments (880 or 2,014 metabolites, respectively, “common” in **Figure 1** and Supplementary Table S5). Conversely, a total of 273 metabolites were identified only in one of the treatments (**Figure 1**), Control (C) being the sample that showed the highest number of characteristic metabolites (78), followed by T5 (44) and R (42). Otherwise, T5 and C were the treatments that shared a greater number of metabolites (106 metabolites; Supplementary Table S5).

The complexity of the metabolome data was reduced by focusing on the specific pathways in relation to the relative abundance of the metabolites identified. Heatmap-clustering analysis (**Figure 2** and Supplementary Figure S1), distinguished four different groups in relation to metabolic pathways identified: C, T1/2-T1, T2-T3, and T5-R group that is highly separated of C and shorter-term treatments.



On the other hand, the clustering based on the profile of KEGG pathways did not show any clear trend, each group presenting a different pathway profile across the sample times, e.g., glucosinate biosynthesis and flavone and flavonol biosynthesis increased their activities in treatments T1/2, T1, T2, or T3; conversely, pentose phosphate pathway and flavonoid biosynthesis showed the maximum in T5 and/or R. Other pathways, such as the biosynthesis of unsaturated fatty acids increased the activity in T2, T3, and T5 or only in the longer-term exposure (T3 and T5) as arginine and proline metabolism pathways. In contrast, pathways related with phenylalanine metabolism enhance their activity during the stress but decrease at T5 (at the levels of C and R). Lastly, riboflavin metabolism showed a substantial increase only in T5 and glutathione metabolism showed a fluctuating tendency, highlighting its maximum in T3.

These results seem to underlie two different types of heat-response in *P. radiata* (shorter-term and longer-term response) where different pathways are necessary at a different time of stress.

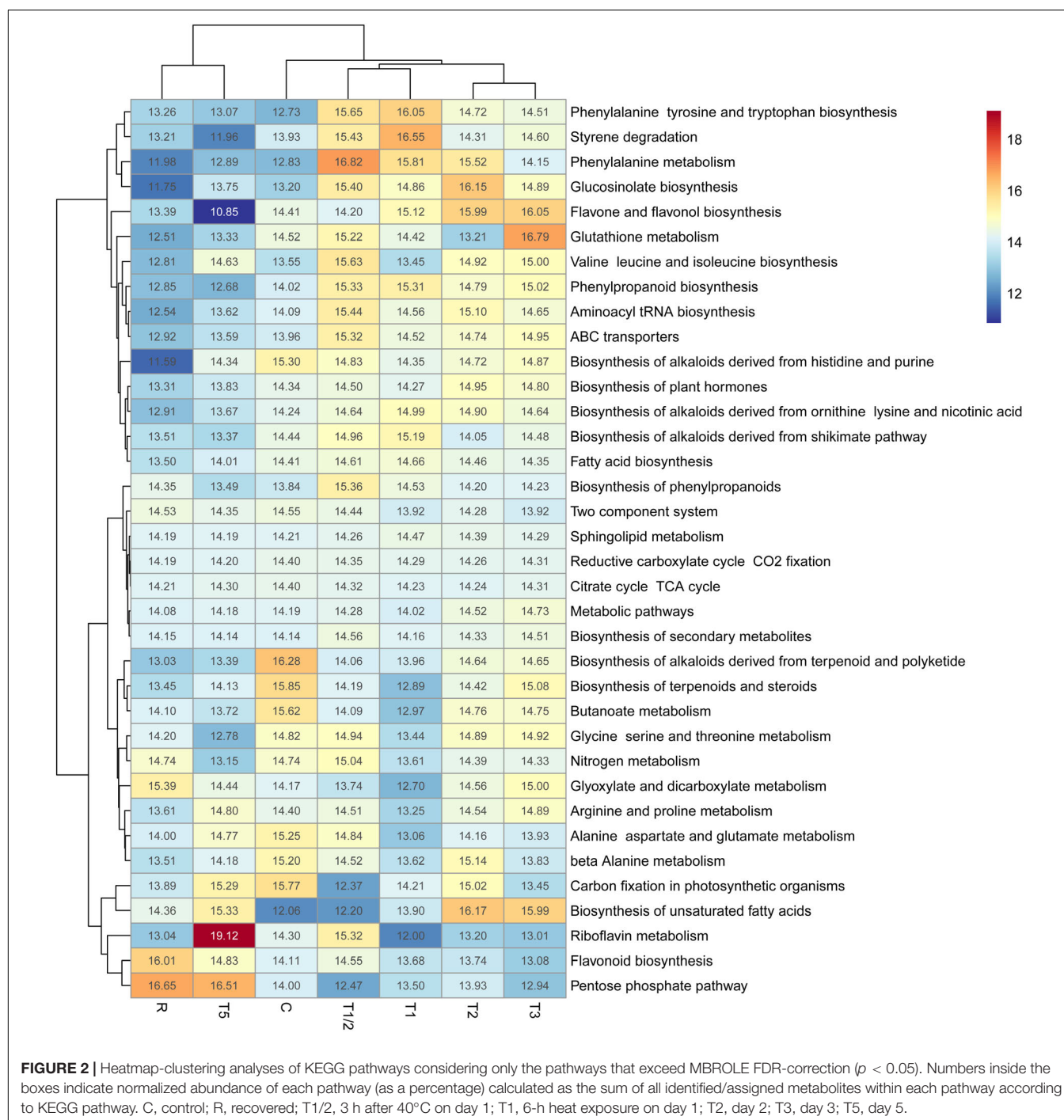
### Shorter-Term and Longer-Term Responses to Heat Stress Were Confirmed by Integrative Analysis of Metabolome and Physiological Datasets

To simplify the dimensionality of the results and integrate metabolome data with physiological datasets, sPLS (Figure 3A and Supplementary Table S6) and PCA (Supplementary Figure S2 and Supplementary Table S7) analyses were used. sPLS and PCA scores revealed a trajectory of the different sampling times by the combination of the two main components. Recovered (R) and long-term exposed (T5) plants were separated from the other treatments by the first component, which seems to be gathering the variance related to long-term stress adaption, while the second component, considering top correlated variables (Supplementary Figure S2 and Supplementary Table S6) is related to heat-response/tolerance. This component revealed differences between Control-Recovered and heat-treatments.

Altogether, multivariate results provided hints about two different responses to stress. First, an initial response to stress (shorter-term response), involves increased ABA and SA activities in samples T1/2, T1, T2, and T3 (Figure 3A, Supplementary Table S6a, Comp 1). Secondly, adaptive mechanisms, such as TSS and ZR, seem to be involved in the distinction between T5 and T3, as well as between the recovered and control plants according to sPLS analysis (Figure 3A, Supplementary Table S6a, Comp 1).

The interaction networks built from this analysis (Figure 3B and Supplementary Table S6) showed a complex correlation between different metabolites and hormones and physiological parameters. The important role of CKs was demonstrated during heat response, as well as the links of these hormones with different compounds of the metabolome and physiological parameters in response to high temperatures. Two main nodes in the constructed network were detected: ZR and iPA nodes. The ZR node was negatively correlated with numerous saturated and unsaturated fatty acids (C16:0 and C18:0 with their unsaturated forms), flavonoids (including kandelin A-1 or eujambolin), amino acids (such as L-proline and L-serine), and L-phenylalanine (key metabolite in several pathways, such as phenylalanine metabolism, phenylpropanoid biosynthesis and biosynthesis of plant hormones). The presence of L-proline in the network indicates the relevance of the accumulation of this amino acid in longer-term response (T3 and T5) to high temperatures in *P. radiata*, as it has already been showed in Escandón et al. (2016) and confirmed by the increased activity of its metabolism (Figure 2, arginine and proline metabolism). Additionally, ZR was positively correlated with fatty acids involved in sphingolipid metabolism, terpenoids (including abscisic-alcohol 11-glucoside) and other secondary metabolites: dihydrokaempferol (related to flavonoid biosynthesis) and *cis*-1,2-dihydro-3-ethylcatechol (implicated in the degradation of aromatic compounds). In this node, it is also important to note that the physiological parameter, EL, appeared negatively correlated by an unknown metabolite (p537) to IAA and DHZR which, in turn, were also positively linked to tetracosanoic acid (saturated fatty acid, C24:0). This confirms the relevance of these hormones (Du et al., 2013;



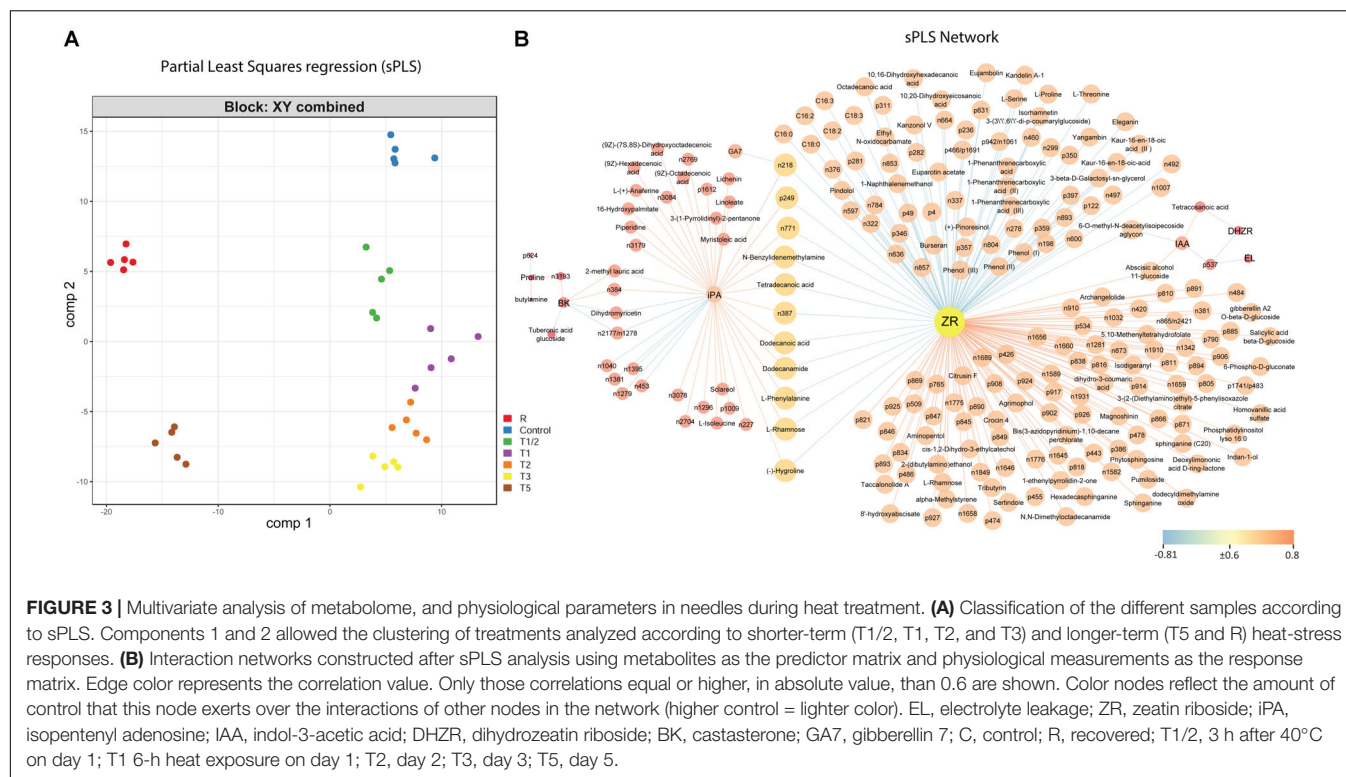


Černý et al., 2014) and fatty acid signaling (Los and Murata, 2004) to repair membrane damage related to heat stress. On the other hand, IPA was positively correlated with fatty acids, such as tetradecanoic acid and (9Z)-octadecenoic acid, as well as metabolites involved in diterpenoid biosynthesis (sclareol); and this hormone was negatively correlated with compounds involved in flavonoid biosynthesis, such as dihydromyricetin.

Interestingly, the analysis of the network dynamics and particularly the quantitation of the represented variables showed

a two-step response, with T3 as the transition point between shorter-term and longer-term responses (Supplementary Movie S1). This observation is consistent with the conducted sPLS and PCA analyses. This behavior can be considered an adaptive mechanism to cope with rapid environmental changes that occur daily (i.e., sun heat following a rainy period). In this case, plants require a mechanism to quickly overcome the first impact of stress; however, the physiology should return to ideal values after the removal of the stress factor





in order to achieve an optimal energetic balance. A previous work (Escandón et al., 2016) showed that photosynthetic state was only slightly affected in the first impact to heat stress (T1/2, T1 and T2), recovered to control levels in T3, and even improved on T5, probably related to the beginning of the acclimation process; while lipids peroxidation analysis showed a slight accumulation of MDA in shorter-term exposures (T1/2, T1 and T2) prior to the activation of acclimation mechanisms. On the other hand, if the stress persists the plant must adapt to the new situation with the cost of reducing its growth and reproductive capacity compared to an ideal situation (Bradford and Hsiao, 1982).

## Using Metabolomics to Explore Possible Thermotolerance Biomarkers

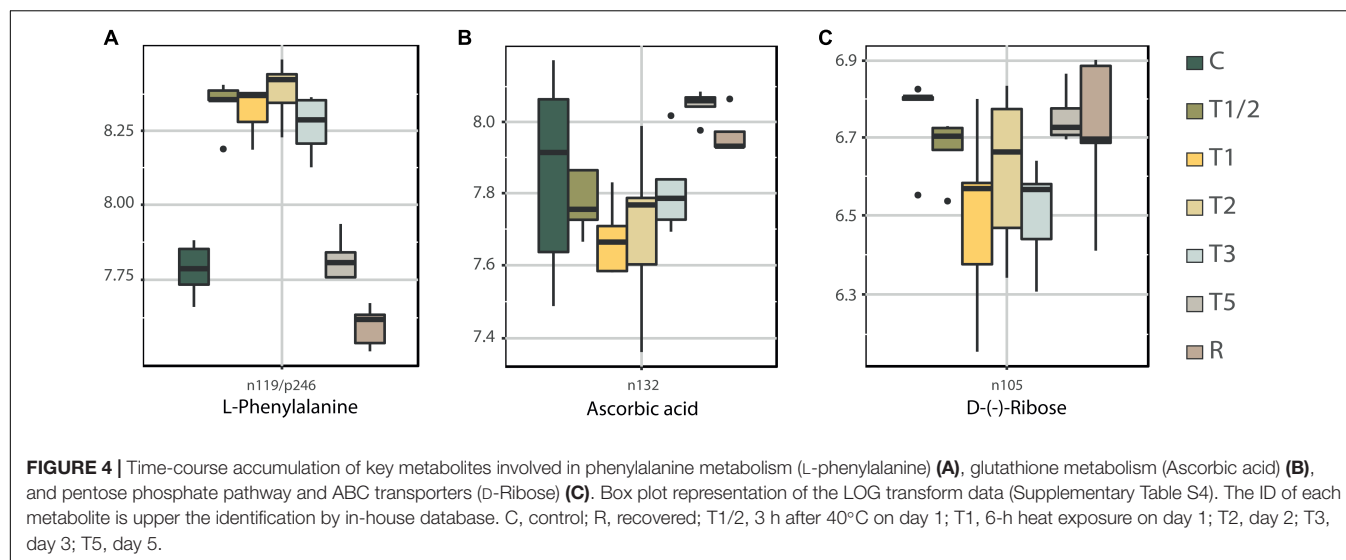
The metabolome analysis revealed the essential role of specific metabolites (Figure 3) and pathways (Figure 2 and Supplementary Figure S1) in relation to high-temperature adaptation in *P. radiata*.

The accumulation of three different metabolites belonging to pathways with higher significant changes across high-temperature treatments is showed in Figure 4. One of these three key compounds is L-phenylalanine which participates in numerous pathways (Meyermans et al., 2000; Wittstock and Halkier, 2000; Teufel et al., 2010; Tzin and Galili, 2010; Yoo et al., 2013), such as phenylalanine metabolism, phenylpropanoid biosynthesis, phenylalanine, tyrosine and tryptophan biosynthesis, glucosinolate biosynthesis, and biosynthesis of plant hormones, most of them highly related/interconnected. Its accumulation profile reflected a level increase from T1/2 to

T3 (Figure 4A), decreasing to control levels in T5 and reaching the lowest accumulation values in R. Phenylpropanoids and flavonoids play a key role in protecting plants against abiotic stress, largely by inhibiting the formation of ROS through a number of different mechanisms (Mierziak et al., 2014). The increment of L-phenylalanine in the shorter treatments may be related to the need to cope with the production of ROS. The decrease in L-phenylalanine in T5 and R could be related to the increased activity of *PHENYLALANINE AMMONIA LYASE* (PAL; crucial enzyme in phenylpropanoid pathway), which transforms L-phenylalanine into *trans*-cinnamic acid (Rivero et al., 2001). The activity increase of this enzyme is considered one of the most important ways of cell acclimation against stress in plants (Levine et al., 1994; Leyva et al., 1995; Rivero et al., 2001).

Other pathway highlighted by heatmap-clustering (Figure 2 and Supplementary Figure S1) was the glutathione metabolism, which showed the highest accumulation in T3. Ascorbic acid is key in this pathway; however, it showed decreased levels in short-term treatments (Figure 4B), reaching the highest values in T5 and R. Ascorbic acid and glutathione are both antioxidants, which are crucial for plant defense against oxidative stress (Noctor and Foyer, 1998). The decrease of ascorbic acid in shorter treatments could be explained by their use in this defense against oxidative damage.

D(–)-ribose (Figure 4C), which participates in pentose phosphate and ABC transporters pathways, showed a decline in shorter-term, recovering the control values in T5 and R. This may indicate that plants reduce their metabolism until T3, recovering the activity in T5 and R when the plants are adapted.



Although sugars play an important role against heat stress in many species (Wahid et al., 2007), in *P. radiata* it has been showed that the total amount of soluble sugars tends to decrease in the first moments of stress (Escandón et al., 2016). This may be because plants maintains growth patterns although they are driven by consuming carbohydrates reserves (Mitchell et al., 2013; Escandón et al., 2016).

Under high temperatures, plants alter lipid composition, causing membranes to become more fluid and thus interrupting membrane processes (Falcone et al., 2004). High-temperature-tolerant plants show an increased presence of saturated fatty acids to counteract increased fluidity during heat stress (Grover et al., 2000; Umesha, 2005). Levels of candidate metabolites related to biosynthesis of unsaturated fatty acids pathway are represented in **Figure 5** including both saturated fatty acids (**Figures 5A,D**) and unsaturated fatty acids (**Figures 5B,C,E,F**). Saturated fatty acids, hexadecanoic acid (C16:0, **Figure 5A**) and C18:0 (**Figure 5D**), shared an increasing tendency in short-term response, starting to fall in T3 (in the case of C16:0) or earlier in T2 (C18:0). In both, the lowest accumulation values were reached in R. This saturated fatty acids raise is consistent with the hypothesis by Grover et al. (2000), which indicates the plants increase saturation until seems to be already acclimated to the heat stress.

In the case of unsaturated fatty acids, they showed different patterns depending on the studied fatty acid. C16:2 and C18:2 showed a slight decrease in short-term response (**Figures 5B,E**), conversely, C16:3 and C18:3 (**Figures 5C,F**) presented increased levels in T1/2 and T1. In longer-term response, the most common tendency is the recovery at the control values (**Figures 5B,C,E**) even surpassing them, except for C18:3 where this only occurs in R (**Figure 5F**). This differential accumulation patterns seems to evidence the different roles of each unsaturated fatty acid during heat-stress response.

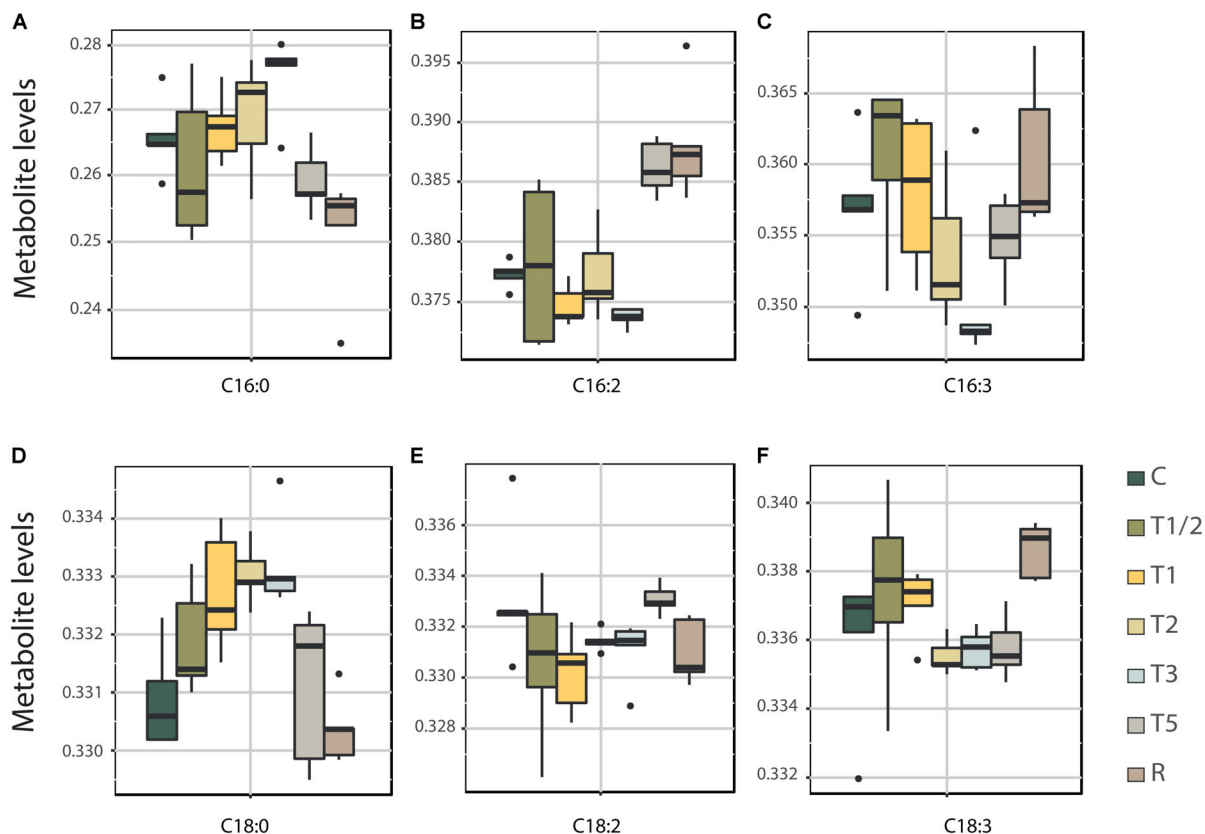
Flavonoids related pathways are also crucial elements according to heatmap-clustering analysis (**Figure 2** and Supplementary Figure S1). Bibliography indicate that flavonoids and anthocyanins are essential compounds to prevent and

protect the plants against different abiotic and biotic stress (Dao et al., 2011; Zhang et al., 2012). Results showed that dihydromyricetin (**Figure 6A**), key metabolite in flavonoid biosynthesis pathway, continuously decreases its levels until T3, to then begin its accumulation, reaching the control values in R. On the contrary, flavonoids eujambolin and kandelin A-1 (**Figures 6B,C**), showed the lowest accumulation in T5 and R. Eujambolin presented a decreasing tendency in shorter treatments, showing a strong decrease in T5 and later in R. Conversely, kandelin A-1 showed an increase in shorter treatments in relation to control, displaying a strong decrease in T5 and R.

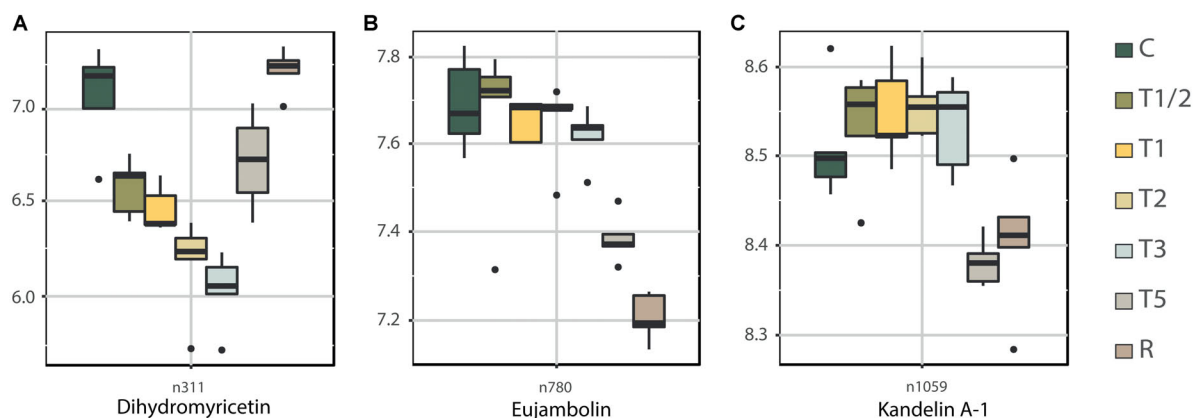
Dihydromyricetin (along with dihydroquercetin and dihydrokaempferol) is a dihydroflavonols implicated in the synthesis of anthocyanidins (Falcone Ferreyra et al., 2012). While, kandelin A-1 is a proanthocyanidin which are synthesized as oligomeric or polymeric end products of one of several branches of the flavonoid pathway, which shares the same upstream pathway with anthocyanins (He et al., 2008). Anthocyanins are usually accumulated during heat stress in vegetative tissues (Wahid and Ghazanfar, 2006) in order to decrease the transpirational losses caused by lower osmotic potential of the leaf (Chalker-Scott, 2002). However, the complexity of its metabolism and the high number of these compounds makes their study difficult.

Gene expression of *PAL* (**Figure 7A**), two *DESATURASE* (*DES*, **Figures 7B,C**) and a central gene involved in anthocyanins and proanthocyanins biosynthesis pathways, *DIHYDROFLAVONOL 4-REDUCTASE* (*DFR*, **Figure 7D**) (Ayabe et al., 2010; Katsu et al., 2017) were analyzed in order to confirm the possible role of these elements as biomarkers and validate the relevance of L-phenylalanine, fatty acid and flavonoid metabolism in the high-temperature response of *P. radiata*.

These data showed that *PAL* (**Figure 7A**) increases its expression in the first moments to stress (T1/2 and T1), returning to the control values in T2 and T3. In T5 and R



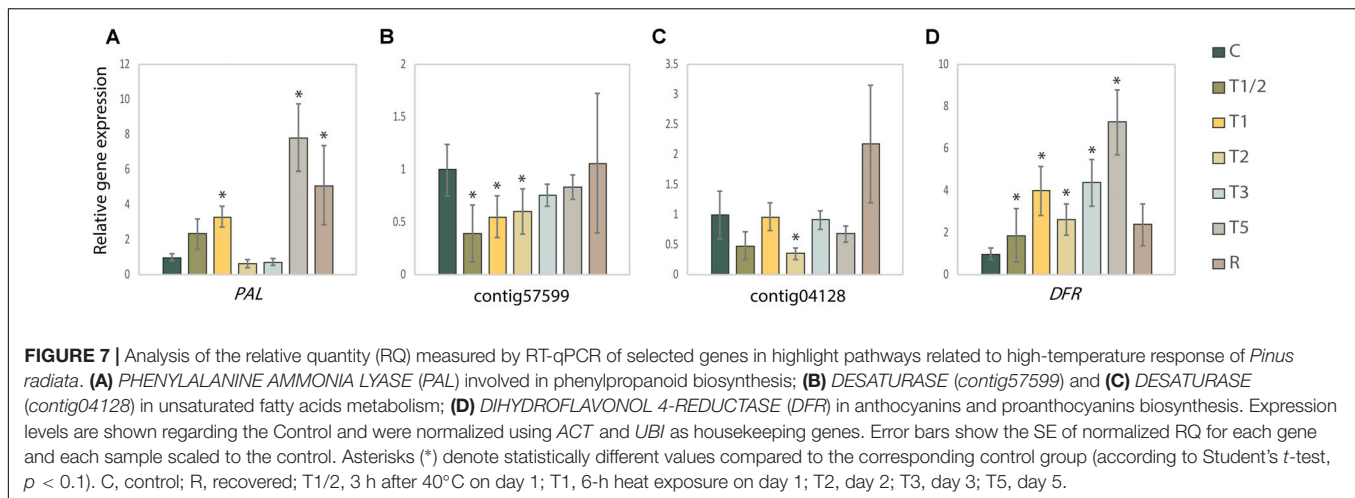
**FIGURE 5 |** Levels of key metabolites of unsaturated fatty acids metabolism across high-temperature treatments. C16: Saturated fatty acid (A) and their unsaturated fatty acids C16:2 (B) and C16:3 (C). C18: Saturated fatty acid (D) and their unsaturated fatty acids C18:2 (E) and C18:3 (F). Box plot representation of LOG transform data for the total content of each fatty acid C16 or C18, respectively. C, control; R, recovered; T1/2, 3 h after 40°C on day 1; T1, 6-h heat exposure on day 1; T2, day 2; T3, day 3; T5, day 5.



**FIGURE 6 |** Levels of key metabolites related to flavonoid biosynthesis across high-temperature treatments: Dihydromyricetin (A), eujambolin (B), and kandelin A-1 (C). Box plot representation of the LOG transform data (Supplementary Table S4). ID of the three metabolite, previously tentatively identified using public database, was validated by the interpretation of their MS/MS spectra (Supplementary Data S1). C, control; R, recovered; T1/2, 3 h after 40°C on day 1; T1, 6-h heat exposure on day 1; T2, day 2; T3, day 3; T5, day 5.

its expression increases drastically, confirming the pattern of decrease of L-phenylalanine (Figure 4A) in these treatments by its consumption. In the case of *DES* (Figures 7B,C), both genes

showed a reduction of the expression in shorter-term treatments: *contig5799* presented the decrease in T1/2, T1 and T2, while *contig04128* only in T2. This reduction is consistent with the



increase of saturated fatty acids identified (C16:0 and C18:0, **Figures 5A,D**). Finally, *DFR* (**Figure 7D**) showed an increase of expression across of stress time, reaching a peak in T5 and return to the control values in R. These results are according dihydromyricetin levels quantified (**Figure 6A**), which seems to be being consumed for the production of anthocyanins in response to high-temperature stress.

## DISCUSSION

### Metabolome Characterization: Dynamic of Metabolism Throughout High-Temperature Stress

The plant metabolome responds to an unfavorable environment in a dynamic way, being favored a characteristic type of metabolic pathways at every moment of stress. Depending on the timing, different compounds can be identified, as stress signal transduction molecules, stress metabolism by-products, or molecules that are part of the plant acclimation response (Shulaev et al., 2008). According to this, the results of this work showed that the response of *P. radiata* at high temperatures activates the synthesis of specific metabolites at each time of the stress (**Figure 1**) showing every sample-time a significant number of unique metabolites. Additionally, these results seem to reveal the beginning of the acclimation process in T5 that showed the greatest number of unique metabolites but also the greatest number of shared metabolites with the control plants (106 metabolites, Supplementary Table S4).

In-depth analysis of metabolite accumulation in relation to KEGG pathways (**Figure 2** and Supplementary Figure S1) allowed to confirm the high dynamism of the metabolome in relation to high-temperature response (Escandón et al., 2017). Although two main trends in the KEGG pathways have been identified by heatmap-clustering (**Figure 2** and Supplementary Figure S1), each pathway seem to have a specific role in a particular moment of the stress, seemingly following an orchestrated succession, particularly, in the case of flavonoids

related pathways. Thus, phenylpropanoid biosynthesis seems to be activated in the first contact with stress (T1/2-T1), while in T2-T3 the most active route is flavone and flavonol biosynthesis, to finally increase the activity of the general pathway of flavonoid biosynthesis in T5 and R. The importance of the flavonoids across the heat stress is evident. According to bibliography the compounds in this group act as antioxidants, contributing to the adaptation to environmental changes such as cold, high temperatures or irradiation (Gould et al., 2002; Doerfler et al., 2013, 2014; Meijón et al., 2016).

Fatty acids also play a fundamental role in the response to stress in *P. radiata*. Biosynthesis of unsaturated fatty acids is a very active pathway in the longer exposure treatments (T2, T3 and T5), while saturated fatty acid biosynthesis is more relevant in the first moments of exposure to heat (T1/2 and T1). A reduction of unsaturated fatty acid content and increase in saturated fatty acids content has been positively associated with heat tolerance as they counteract the increase in membrane fluidity caused by high temperature (Grover et al., 2000; Larkindale and Huang, 2004). However, this hypothesis has not been fully confirmed since that unsaturated fatty acid is also known to be key elements in heat stress signaling (Königshofer et al., 2008) and to provide other essential characteristics to lipid membrane (Falcone et al., 2004), showing an increase in its accumulation at high temperatures.

In the case of *P. radiata*, biosynthesis of fatty acid is a key mechanism to overcome the fluidization of the membrane until T2 at the beginning of the stress, when the plants had overcome this fluidization. The global increment in saturated fatty acid would be consequence of the increase of the amount of a specific saturated fatty acid (e.g., hexadecanoic acid) or the accumulation of fatty acids with lower numbers of double bonds (the reduction of C18:3 to C18:2 or C18:1; Wang et al., 2017). From T2, plants activate the biosynthesis of unsaturated fatty acids which may be necessary to stabilize photosynthesis (Gombos et al., 1994) or activate signaling (e.g. lipid or calcium signaling) (Königshofer et al., 2008). These observations support the idea of Falcone et al. (2004) that the unsaturation level of lipid membranes also plays an important role in the plant's ability to tolerate high



temperatures, although other characteristics of plants membrane lipids are also likely to be important.

The study of the metabolome across heat stress and recovery, has given the possibility to deepen in the different dynamics of the metabolomic pathways in *P. radiata*, establishing the existence of key pathways in different moments of the stress response.

## Integrative Analysis of Physiological Response and Metabolome Dynamic

This work revealed that high temperature has a complex impact on cell function, suggesting that many and complex processes are involved in heat-resistance processes. However, the use of system-wide approach and integrative bioinformatics tools have allowed the understanding of the molecular basis, the identification of injury mediators, and the characterization of associated biomarkers. Thus, CKs, fatty acid metabolism and flavonoid and terpenoid biosynthesis were revealed as the most relevant pathways related to shorter-term and longer-term response clusters confirmed by multivariate analysis (**Figure 3A**), being ZR and iPA the key hormones that coordinate the different response processes (**Figure 3B**).

Some of the secondary metabolites with a higher loading in the shorter-term and longer-term response cluster included dihydrokaempferol, *cis*-1,2-dihydro-3-ethylcatechol and dihydromyricetin, which are all crucial elements in flavonoid metabolism. Flavonoids are a biologically and chemically diverse group widely represented in plants. Their diversity and multi-functionality demonstrate their importance in plants. Terpenoids are the other significant group of secondary metabolites identified in relation to the clusters, which are produced by a variety of plants and particularly in conifers (Zulak and Bohlmann, 2010). These volatile compounds are emitted by plants and play an important role in the interaction with their environment (Tholl, 2015). The best known and most studied group of terpenoids is the sesquiterpenoid plant hormone ABA, the central element in the plant stress response (McCourt et al., 2005; Raghavendra et al., 2010). Absciscic alcohol 11-glucoside, the glycosylated form of ABA, was also positively correlated to ZR as a key element in the network (**Figure 3B**). Both groups of secondary metabolites, flavonoids and terpenoids, are important in plant growth, development, and response against biotic and abiotic stress (Alder et al., 2012), as well as in plant adaptation to variable environmental conditions (Kliebenstein, 2004). Plasma membrane fluidity has been described to be an important temperature sensor in plants that appears to lie upstream of the unfolded protein response (Wu et al., 2012). Increased membrane fluidity appears to open calcium channels in the plasma membrane and the resultant inflow of calcium triggers signaling cascades, including an H<sub>2</sub>O<sub>2</sub> burst (Königshofer et al., 2008), which activates the heat-stress response. In fact, fatty acids are the most prominent group of compounds revealed in the heat-stress response network, after CKs.

Integrated physiological and metabolome analysis across high-temperature treatments (Supplementary Movie S1) in *P. radiata* emphasizes the complex dynamics of the metabolome in response to heat stress and suggests the existence of a turning point (T3) at which *P. radiata* plants changed from an

initial stress response (shorter-term response) to an acclimation one (longer-term response). The video highlights how the metabolites that are positively related to ZR increase drastically its accumulation in T5 and R. This is regulated by a complex interaction network that involves multiple pathways and groups of compounds where possible biomarkers of thermotolerance processes could be found.

## Evaluation of Proposed Thermotolerance Biomarkers in *P. radiata*

In plants, the concept of biomarker could be defined as “a characteristic that is objectively measured or evaluated as a predictor of plant performance” (Fernandez et al., 2016). The use of biomarkers originated from the field of medicine, but in plants in recent years, many authors have used metabolites as indicators for estimating plant performance under stress conditions (Quistian et al., 2011; Degenkolbe et al., 2013; Nam et al., 2015; Obata et al., 2015). One of main goals in this study is to find metabolic markers of high-temperature tolerance, potentially useful for *P. radiata* breeding programs.

L-phenylalanine, hexadecanoic acid (C16:0), and dihydromyricetin were confirmed as the three strongest biomarkers between the proposed candidates related to the results of this work. Their biological relevance, statistical strength in integrative analysis, and accumulation profile across the samples times analyzed validate their future use as thermotolerance biomarkers in *P. radiata*.

L-phenylalanine seems to be one of the clearest candidates. L-phenylalanine is crucial in numerous KEGG pathways which changed significantly during stress, particularly in shorter-term treatments (e.g., phenylalanine metabolism, phenylpropanoid biosynthesis, phenylalanine, tyrosine and tryptophan biosynthesis, glucosinolate biosynthesis, and biosynthesis of plant hormones). In addition, integrative analysis of metabolome and physiological measurements showed its relevance as a key compound in the network related to high-temperature response (**Figure 3B**). L-phenylalanine showed a direct correlation with the CKs associated with the trigger of response (negatively with ZR and positively with iPA). Old studies (Dedio and Clark, 1971; Deikman and Hammer, 1995) have already seen how the application of CKs stimulated flavonoids pathways in which L-phenylalanine is involved (flavonoids are synthesized by the phenylpropanoid metabolic pathway), such as the production of isoflavone and anthocyanins synthesis. More recently, Angelova et al. (2001) and Ali and Abbas (2003) have debated the impact of CKs on the accumulation of flavonol glycosides and Hamayun et al. (2015) found out how kinetin modulates isoflavone contents under salinity stress. Furthermore, *PAL* expression analysis (**Figure 7A**) confirms the beginning of the acclimation process in T5 when L-phenylalanine levels (**Figure 4A**) showed a significant decrease, validating the high value of this metabolite as a biomarker.

The second candidate biomarker was chosen within saturated fatty acids, given their importance in stabilizing membrane fluidity during heat stress (Grover et al., 2000; Larkindale and Huang, 2004; Wang et al., 2017). Hexadecanoic acid (also known as palmitic acid or C16:0) increases its accumulation in shorter

heat exposures, suffering a decrease in T5 and R, when the plant can be already acclimatized to stress (**Figure 5E**). Moreover, the low expression levels of both *DES* (**Figures 7B,C**) in shorter treatments underline the importance of saturated fatty acids, such as hexadecanoic acid, in the early response to high temperatures. This saturated fatty acid has already been studied by other authors such as Alfonso et al. (2001), Falcone et al. (2004), and Larkindale and Huang (2004) in relation to high-temperatures response. It has been even observed that the thermotolerance of an *Arabidopsis* mutant deficient in palmitic acid unsaturation is enhanced (Kunst et al., 1989). Furthermore, hexadecanoic is showed as an important element by the performed integrative analysis, appearing interconnected with the master regulator ZR in the network.

The last strong biomarker proposed is dihydromyricetin, also known as ampelopsin. Dihydromyricetin is a flavanonol included in anthocyanin biosynthesis inside flavonoid biosynthesis pathway. Flavonoids and anthocyanins have a photoprotective and antioxidant role (Dao et al., 2011; Zhang et al., 2012). Dihydromyricetin is consumed during anthocyanins synthesis, reducing its accumulation in shorter treatments and reaching the lowest levels in T3 (**Figure 6A**). This is validated by the increased expression of *DFR* across the exposure to high temperatures (**Figure 7D**), showing its maximum expression in T5 when dihydromyricetin begins to be accumulated again. Anthocyanins are produced under a variety of stresses such as UV-B (Valledor et al., 2012), low temperatures (Krol et al., 1995), or salinity (Wahid and Ghazanfar, 2006), although only a few studies have dealt with the effect of high temperatures on anthocyanin accumulation (Wahid and Close, 2007; Correia et al., 2014; de Leonardis et al., 2015). Dihydromyricetin was also prominent in multivariate and integrative analysis showing a negative correlation with iPA and BK. Cytokinin increases anthocyanin content and the transcript levels of *PRODUCTION OF ANTHOCYANIN PIGMENT 1* (Das et al., 2012) which is according to the dropping levels of dihydromyricetin (anthocyanin precursor) observed in the first impact of the heat stress. Among many other biological functions, anthocyanins are considered the first line of defense against oxidative stress (Gould et al., 2002), scavenging oxygen radicals, and inhibiting lipid peroxidation (Chalker-Scott, 2002; Ling et al., 2007).

## CONCLUSION

This work shows that high temperatures induced a quick and dynamic change in the metabolome of *P. radiata*, in order to maintain homeostasis and facilitate survival. Integrative study of metabolome across high-temperature exposure and recovery plants allowed reaching a global view of molecular mechanism behind high-temperature response in *P. radiata*,

revealing complex interaction networks that involve CKs, fatty acid metabolism, and flavonoid and terpenoid biosynthesis being ZR and iPA the master regulators that trigger the global response. Additionally, novel potential thermotolerance biomarkers such as L-phenylalanine, hexadecanoic acid and dihydromyricetin have been proposed. However, these potential biomarkers need to be validated in further studies in *P. radiata* and their possible universality analyzed in other species.

## AUTHOR CONTRIBUTIONS

ME, MM, MC, and LV designed and performed the experimental work. ME processed the samples, integrated the datasets, and completed the metabolome analyses. LV and MM performed the mass spectrometry. GP realized the physiological analysis. ME and JP were involved in the preparation of samples. LV and JP helped with statistical analyses. ME and MM wrote the manuscript while MC, LV, JP, and GP supervised the manuscript. All authors read and approved the final manuscript.

## FUNDING

This publication is an output of the projects financed by the Spanish Ministry of Economy, Industry and Competitiveness (AGL2014-54995-P and AGL2016-77633-P), the Government of Principado de Asturias (GRUPIN14-055), FEDER funding through COMPETE (Project: UID/AMB/50017/2013), and by National Funds through the Portuguese Foundation for Science and Technology (FCT) within the Project PTDC/AGR-FOR/2768/2014. ME was supported by a fellowship from the Severo Ochoa Program (BP11117; Government of Principado de Asturias, Spain). JP was supported by a fellowship from the FPU (AP2010-5857; Ministry of Education, Spain). FCT (Fundação para a Ciência e a Tecnologia, Portugal) supported the fellowships of GP (SFRH/BPD/101669/2014). The Spanish Ministry of Economy and Competitiveness supported MM and LV through the Ramón y Cajal program (RYC-2014-14981 and RYC-2015-17871, respectively). FCT/MEC (Portugal) co-funding by the FEDER, within the PT2020 Partnership Agreement and Compete 2020 provide financial support to Centre for Environmental and Marine Studies (CESAM – UID/AMB/50017).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.00485/full#supplementary-material>

## REFERENCES

- Acharya, B. R., and Assmann, S. M. (2009). Hormone interactions in stomatal function. *Plant Mol. Biol.* 69, 451–462. doi: 10.1007/s11103-008-9427-0
- Alder, A., Jamil, M., Marzorati, M., Bruno, M., Vermathen, M., Bigler, P., et al. (2012). The path from beta-carotene to carlactone, a strigolactone-like plant hormone. *Science* 335, 1348–1351. doi: 10.1126/science.1218094
- Alfonso, M., Yruela, I., Almarcegui, S., Torrado, E., Pérez, A. M., and Picorel, R. (2001). Unusual tolerance to high temperatures in a new herbicide-resistant

- D1 mutant from *Glycine max* (L.) Merr. cell cultures deficient in fatty acid desaturation. *Planta* 212, 573–582. doi: 10.1007/s004250000421
- Ali, R. M., and Abbas, H. M. (2003). Response of salt stressed barley seedlings to phenylurea. *Plant Soil Environ.* 49, 158–162. doi: 10.17221/4107-PSE
- Angelova, Y., Petkova, S., Zozikova, E., Kotseva, E., and Iliev, L. (2001). Effects of kinetin and 4PU-30 on the growth and the content of polyphenols in tobacco callus tissue. *Bulg. J. Plant Physiol.* 27, 36–42.
- Ayabe, S., Uchiyama, H., Aoki, T., and Akashi, T. (2010). 1.24 - plant phenolics: phenylpropanoids. *Compr. Nat. Prod.* 1, 929–976. doi: 10.1016/B978-008045382-8.00023-X
- Bradford, K. J., and Hsiao, T. C. (1982). “Physiological responses to moderate water stress,” in *Physiological Plant Ecology II. Water Relations and Carbon Assimilation. Encyclopedia of Plant Physiology*, eds O. Lange, P. Nobel, C. Osmond, and H. Ziegler (Berlin: Springer-Verlag), 263–324.
- Bruggeman, F. J., and Westerhoff, H. V. (2007). The nature of systems biology. *Trends Microbiol.* 15, 45–50. doi: 10.1016/j.tim.2006.11.003
- Černý, M., Jedelsky, P. L., Novak, J., Schlosser, A., and Brzobohaty, B. (2014). Cytokinin modulates proteomic, transcriptomic and growth responses to temperature shocks in *Arabidopsis*. *Plant Cell Environ.* 37, 1641–1655. doi: 10.1111/pce.12270
- Chalker-Scott, L. (2002). Do anthocyanins function as osmoregulators in leaf tissues? *Adv. Bot. Res.* 37, 103–127. doi: 10.1016/S0065-2296(02)37046-0
- Cline, M. S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., et al. (2007). Integration of biological networks and gene expression data using Cytoscape. *Nat. Protoc.* 2, 2366–2382. doi: 10.1038/nprot.2007.324
- Correia, B., Luis, J., Valledor, L., Almeida, T., Santos, C., Cañal, M. J., et al. (2014). Analysis of the expression of putative heat-stress related genes in relation to thermotolerance of cork oak. *J. Plant Physiol.* 171, 399–406. doi: 10.1016/j.jplph.2013.12.004
- Dao, T. T. H., Linthorst, H. J. M., and Verpoorte, R. (2011). Chalcone synthase and its functions in plant resistance. *Phytochem. Rev.* 10, 397–412. doi: 10.1007/s11101-011-9211-7
- Das, P. K., Shin, D. H., Choi, S.-B., Yoo, S.-D., Choi, G., and Park, Y.-I. (2012). Cytokinins enhance sugar-induced anthocyanin biosynthesis in *Arabidopsis*. *Mol. Cells* 34, 93–101. doi: 10.1007/s10059-012-0114-2
- de Leonardi, A. M., Fragasso, M., Beleggia, R., Ficco, D. B. M., de Vita, P., and Mastrangelo, A. M. (2015). Effects of heat stress on metabolite accumulation and composition, and nutritional properties of durum wheat grain. *Int. J. Mol. Sci.* 16, 30382–30404. doi: 10.3390/ijms161226241
- Dedio, W., and Clark, K. W. (1971). Influence of cytokinins on isoflavone and anthocyanin synthesis in red clover seedlings. *Pestic. Sci.* 2, 65–68. doi: 10.1002/ps.2780020204
- Degenkolbe, T., Do, P. T., Kopka, J., Zuther, E., Hinch, D. K., and Köhl, K. I. (2013). Identification of drought tolerance markers in a diverse population of rice cultivars by expression and metabolite profiling. *PLoS One* 8:e63637. doi: 10.1371/journal.pone.0063637
- Deikman, J., and Hammer, P. E. (1995). Induction of anthocyanin accumulation by cytokinins in *Arabidopsis thaliana*. *Plant Physiol.* 108, 47–57. doi: 10.1104/pp.108.1.47
- Doerfler, H., Lyon, D., Nagele, T., Sun, X., Fragner, L., Hadacek, F., et al. (2013). Granger causality in integrated GC-MS and LC-MS metabolomics data reveals the interface of primary and secondary metabolism. *Metabolomics* 9, 564–574. doi: 10.1007/s11306-012-0470-0
- Doerfler, H., Sun, X., Wang, L., Engelmeier, D., Lyon, D., and Weckwerth, W. (2014). mzGroupAnalyzer—predicting pathways and novel chemical structures from untargeted high-throughput metabolomics data. *PLoS One* 9:e96188. doi: 10.1371/journal.pone.0096188
- Du, H., Liu, H., and Xiong, L. (2013). Endogenous auxin and jasmonic acid levels are differentially modulated by abiotic stresses in rice. *Front. Plant Sci.* 4:397. doi: 10.3389/fpls.2013.00397
- EEA (2012). *Climate Change, Impacts and Vulnerability in Europe*. Available at: [http://www.eea.europa.eu/publications/climate-impacts-and-vulnerability-2012/at\\_download/file](http://www.eea.europa.eu/publications/climate-impacts-and-vulnerability-2012/at_download/file)
- Escandón, M., Cañal, M. J., Pascual, J., Pinto, G., Correia, B., Amaral, J., et al. (2016). Integrated physiological and hormonal profile of heat-induced thermotolerance in *Pinus radiata*. *Tree Physiol.* 36, 63–77. doi: 10.1093/treephys/tpv127
- Escandón, M., Valledor, L., Pascual, J., Pinto, G., Cañal, M., and Meijón, M. (2017). System-wide analysis of short-term response to high temperature in *Pinus radiata*. *J. Exp. Bot.* 68, 3629–3641. doi: 10.1093/jxb/erx198
- Falcone, D. L., Ogas, J. P., and Somerville, C. R. (2004). Regulation of membrane fatty acid composition by temperature in mutants of *Arabidopsis* with alterations in membrane lipid composition. *BMC Plant Biol.* 4:17. doi: 10.1186/1471-2229-4-17
- Falcone Ferreyra, M. L., Rius, S. P., and Casati, P. (2012). Flavonoids: biosynthesis, biological functions, and biotechnological applications. *Front. Plant Sci.* 3:222. doi: 10.3389/fpls.2012.00222
- Fernández, de Simón, B., Sanz, M., Cervera, M. T., Pinto, E., Aranda, I., et al. (2017). Leaf metabolic response to water deficit in *Pinus pinaster* Ait. relies upon ontogeny and genotype. *Environ. Exp. Bot.* 140, 41–55. doi: 10.1016/j.envexpbot.2017.05.017
- Fernandez, O., Urrutia, M., Bernillon, S., Giauffret, C., Tardieu, F., Le Gouis, J., et al. (2016). Fortune telling: metabolic markers of plant performance. *Metabolomics* 12:158. doi: 10.1007/s11306-016-1099-1
- Gibbons, H., O’Gorman, A., and Brennan, L. (2015). Metabolomics as a tool in nutritional research. *Curr. Opin. Lipidol.* 26, 30–34.
- Gombos, Z., Wada, H., Hideg, E., and Murata, N. (1994). The unsaturation of membrane lipids stabilizes photosynthesis against heat stress. *Plant Physiol.* 104, 563–567. doi: 10.1104/pp.104.2.563
- Gould, K. S., McKelvie, J., and Markham, K. R. (2002). Do anthocyanins function as antioxidants in leaves? Imaging of H<sub>2</sub>O<sub>2</sub> in red and green leaves after mechanical injury. *Plant Cell Environ.* 25, 1261–1269. doi: 10.1046/j.1365-3040.2002.00905.x
- Grover, A., Agarwal, M., Katiyar-Agarwal, S., Sahi, C., and Agarwal, S. (2000). Production of high temperature tolerant transgenic plants through manipulation of membrane lipids. *Curr. Sci.* 79, 557–559.
- Hamayun, M., Hussain, A., Khan, S. A., Irshad, M., Khan, A. L., Waqas, M., et al. (2015). Kinetin modulates physio-hormonal attributes and isoflavone contents of Soybean grown under salinity stress. *Front. Plant Sci.* 6:377. doi: 10.3389/fpls.2015.00377
- He, F., Pan, Q.-H., Shi, Y., and Duan, C.-Q. (2008). Biosynthesis and genetic regulation of proanthocyanidins in plants. *Molecules* 13, 2674–2703. doi: 10.3390/molecules13102674
- Hellemans, J., Mortier, G., Paepe, A., De, Speleman, F., and Vandesompele, J. (2007). qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol.* 8:R19. doi: 10.1186/gb-2007-8-2-r19
- Hummel, J., Niemann, M., Wienkoop, S., Schulze, W., Steinhauser, D., Selbig, J., et al. (2007). ProMEX : a mass spectral reference database for proteins and protein phosphorylation sites. *BMC Bioinformatics* 8:216. doi: 10.1186/1471-2105-8-216
- Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* 40, D109–D114. doi: 10.1093/nar/gkr988
- Katsu, K., Suzuki, R., Tsuchiya, W., Inagaki, N., Yamazaki, T., Hisano, T., et al. (2017). A new buckwheat dihydroflavonol 4-reductase (DFR), with a unique substrate binding structure, has altered substrate specificity. *BMC Plant Biol.* 17:239. doi: 10.1186/s12870-017-1200-6
- Kliebenstein, D. J. (2004). Secondary metabolites and plant/environment interactions: a view through *Arabidopsis thaliana* tinted glasses. *Plant Cell Environ.* 27, 675–684. doi: 10.1111/j.1365-3040.2004.01180.x
- Kolde, R. (2015). *heatmap: Pretty Heatmaps. R package version 0.7.7*. Available at: <https://cran.r-project.org/web/packages/heatmap/>
- Königshofer, H., Trombala, H., and Löppert, H. (2008). Early events in signalling high-temperature stress in tobacco BY2 cells involve alterations in membrane fluidity and enhanced hydrogen peroxide production. *Plant Cell Environ.* 31, 1771–1780. doi: 10.1111/j.1365-3040.2008.01880.x
- Krol, M., Gray, G. R., Huner, N. P. A., Hurry, V. M., Öquist, G., and Malek, L. (1995). Low-temperature stress and photoperiod affect an increased tolerance to photoinhibition in *Pinus banksiana* seedlings. *Can. J. Bot.* 73, 1119–1127. doi: 10.1139/b95-122
- Kunst, L., Browse, J., and Somerville, C. (1989). A mutant of *Arabidopsis* deficient in desaturation of palmitic acid in leaf lipids. *Plant Physiol.* 90, 943–947.



- Larkindale, J., and Huang, B. (2004). Thermotolerance and antioxidant systems in *Agrostis stolonifera*: involvement of salicylic acid, abscisic acid, calcium, hydrogen peroxide, and ethylene. *J. Plant Physiol.* 161, 405–413. doi: 10.1078/0176-1617-01239
- Lê Cao, K.-A., Gonzalez, I., and Dejean, S. (2009). integrOmics: an R package to unravel relationships between two omics data sets. *Bioinformatics* 25, 2855–2856. doi: 10.1093/bioinformatics/btp515
- Levine, A., Tenhaken, R., Dixon, R., and Lamb, C. (1994). H<sub>2</sub>O<sub>2</sub> from the oxidative burst orchestrates the plant hypersensitive disease resistance response. *Cell* 79, 583–593. doi: 10.1016/0092-8674(94)90544-4
- Leyva, A., Jarillo, J. A., Salinas, J., and Martinez-Zapater, J. M. (1995). Low temperature induces the accumulation of phenylalanine ammonia-lyase and chalcone synthase mRNAs of *Arabidopsis thaliana* in a light-dependent manner. *Plant Physiol.* 108, 39–46. doi: 10.1104/pp.108.1.39
- Ling, S., Shu, Z., Sun, S.-L., Peng, C., Wang, X.-J., and Lin, Z.-F. (2007). Antioxidation of anthocyanins in photosynthesis under high temperature stress. *J. Integr. Plant Biol.* 49, 1341–1351. doi: 10.1111/j.1744-7909.2007.00527.x
- López-Ibáñez, J., Pazos, F., and Chagoyen, M. (2016). MBROLE 2.0 – functional enrichment of chemical compounds. *Nucleic Acids Res.* 44, 201–204. doi: 10.1093/nar/gkw253
- Los, D. A., and Murata, N. (2004). Membrane fluidity and its roles in the perception of environmental signals. *Biochim. Biophys. Acta* 1666, 142–157. doi: 10.1016/j.bbame.2004.08.002
- Martinich, J., Crimmins, A., Beach, R. H., Thomson, A., and McFarland, J. (2017). Focus on agriculture and forestry benefits of reducing climate change impacts. *Environ. Res. Lett.* 12:060301.
- McCourt, P., Lumba, S., Tsuchiya, Y., and Gazzarrini, S. (2005). Crosstalk and abscisic acid: the roles of terpenoid hormones in coordinating development. *Physiol. Plant.* 123, 147–152. doi: 10.1111/j.1399-3054.2005.00457.x
- Meijón, M., Feito, I., Oravec, M., Delatorre, C., Weckwerth, W., Majada, J., et al. (2016). Exploring natural variation of *Pinus pinaster* Aiton using metabolomics: is it possible to identify the region of origin of a pine from its metabolites? *Mol. Ecol.* 25, 959–976. doi: 10.1111/mec.13525
- Meyermans, H., Morreel, K., Lapierre, C., Pollet, B., De Bruyn, A., Busson, R., et al. (2000). Modifications in lignin and accumulation of phenolic glucosides in poplar xylem upon down-regulation of caffeoyl-coenzyme A O-methyltransferase, an enzyme involved in lignin biosynthesis. *J. Biol. Chem.* 275, 36899–36909. doi: 10.1074/jbc.M006915200
- Mierziak, J., Kostyn, K., and Kulma, A. (2014). Flavonoids as important molecules of plant interactions with the environment. *Molecules* 19, 16240–16265. doi: 10.3390/molecules191016240
- Mitchell, P. J., O'Grady, A. P., Tissue, D. T., White, D. A., Ottenschlaeger, M. L., and Pinkard, E. A. (2013). Drought response strategies define the relative contributions of hydraulic dysfunction and carbohydrate depletion during tree mortality. *New Phytol.* 197, 862–872. doi: 10.1111/nph.12064
- Monteiro, M. S., Carvalho, M., Bastos, M. L., Guedes, and de Pinho, P. (2013). Metabolomics analysis for biomarker discovery: advances and challenges. *Curr. Med. Chem.* 20, 257–271. doi: 10.2174/092986713804806621
- Nam, M. H., Bang, E., Kwon, T. Y., Kim, Y., Kim, E. H., Cho, K., et al. (2015). Metabolite profiling of diverse rice germplasm and identification of conserved metabolic markers of rice roots in response to long-term mild salinity stress. *Int. J. Mol. Sci.* 16, 21959–21974. doi: 10.3390/ijms160921959
- Noctor, G., and Foyer, C. H. (1998). ASCORBATE and glutathione: keeping active oxygen under control. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49, 249–279. doi: 10.1146/annurev.arplant.49.1.249
- Obata, T., Witt, S., Lisec, J., Palacios-Rojas, N., Florez-Sarasa, I., Yousfi, S., et al. (2015). Metabolite profiles of maize leaves in drought, heat, and combined stress field trials reveal the relationship between metabolism and grain yield. *Plant Physiol.* 169, 2665–2683. doi: 10.1104/pp.15.01164
- Pascual, J., Cañal, M. J., Escandón, M., Meijón, M., Weckwerth, W., and Valledor, L. (2017). Integrated physiological, proteomic, and metabolomic analysis of ultra violet (UV) stress responses and adaptation mechanisms in *Pinus radiata*. *Mol. Cell Proteomics* 16, 485–501. doi: 10.1074/mcp.M116.059436
- Patel, S., and Ahmed, S. (2015). Emerging field of metabolomics: big promise for cancer biomarker identification and drug discovery. *J. Pharm. Biomed. Anal.* 107, 63–74. doi: 10.1016/j.jpba.2014.12.020
- Peng, B., Li, H., and Peng, X.-X. (2015). Functional metabolomics: from biomarker discovery to metabolome reprogramming. *Protein Cell* 6, 628–637. doi: 10.1007/s13238-015-0185-x
- Pluskal, T., Castillo, S., Villar-briones, A., and Orešič, M. (2010). MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* 11:395. doi: 10.1186/1471-2105-11-395
- Quistian, D., Estrada-Luna, A. A., Altamirano-Hernandez, J., Peña Cabrales, J., Montes, de Oca-Luna, R., et al. (2011). Use of trehalose metabolism as a biochemical marker in rice breeding. *Mol. Breed.* 30, 469–477. doi: 10.1007/s11032-011-9636-0
- R Development Core Team (2015). *A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Raghavendra, A. S., Gonugunta, V. K., Christmann, A., and Grill, E. (2010). ABA perception and signalling. *Trends Plant Sci.* 15, 395–401. doi: 10.1016/j.tplants.2010.04.006
- Rivero, R. M., Ruiz, J. M., García, P. C., López-Lefebvre, L. R., Sanchez, E., and Romero, L. (2001). Resistance to cold and heat stress: accumulation of phenolic compounds in tomato and watermelon plants. *Plant Sci.* 160, 315–321. doi: 10.1016/S0168-9452(00)00395-2
- Routaboul, J.-M., Dubos, C., Beck, G., Marquis, C., Bidzinski, P., Loudet, O., et al. (2012). Metabolite profiling and quantitative genetics of natural variation for flavonoids in *Arabidopsis*. *J. Exp. Bot.* 63, 3749–3764. doi: 10.1093/jxb/ers067
- RStudio Team (2015). *RStudio: Integrated Development Environment for R*. Boston, MA: RStudio.
- Schworer, C., Henne, P. D., and Tinner, W. (2014). A model-data comparison of Holocene timberline changes in the Swiss Alps reveals past and future drivers of mountain forest dynamics. *Glob. Chang. Biol.* 20, 1512–1526. doi: 10.1111/gcb.12456
- Shulaev, V., Cortes, D., Miller, G., and Mittler, R. (2008). Metabolomics for plant stress response. *Physiol. Plant.* 132, 199–208. doi: 10.1111/j.1399-3054.2007.01025.x
- Skalák, J., Martin, Č., Jedelský, P., Dobrá, J., Ge, E., Novák, J., et al. (2016). Stimulation of ipt overexpression as a tool to elucidate the role of cytokinins in high temperature responses of *Arabidopsis thaliana*. *J. Exp. Bot.* 67, 2861–2873. doi: 10.1093/jxb/erw129
- Teufel, R., Mascaraque, V., Ismail, W., Voss, M., Perera, J., Eisenreich, W., et al. (2010). Bacterial phenylalanine and phenylacetate catabolic pathway revealed. *Proc. Natl. Acad. Sci. U.S.A.* 107, 14390–14395. doi: 10.1073/pnas.1005399107
- Tholl, D. (2015). Biosynthesis and biological functions of terpenoids in plants. *Adv. Biochem. Eng. Biotechnol.* 148, 63–106. doi: 10.1007/10\_2014\_295
- Tzin, V., and Galili, G. (2010). The biosynthetic pathways for Shikimate and aromatic amino acids in *Arabidopsis thaliana*. *Arabidopsis Book* 8, e0132. doi: 10.1199/tab.0132
- Umesha, S. (2005). *Plant Biotechnology*. New Delhi: TERI Publishing House.
- Valledor, L., Canal, M. J., Pascual, J., Rodriguez, R., and Meijon, M. (2012). Early induced protein 1 (PrELIP1) and other photosynthetic, stress and epigenetic regulation genes are involved in *Pinus radiata* D. don UV-B radiation response. *Physiol. Plant.* 146, 308–320. doi: 10.1111/j.1399-3054.2012.01629.x
- Valledor, L., Escandón, M., Meijón, M., Nukarinen, E., Cañal, M. J., and Weckwerth, W. (2014a). A universal protocol for the combined isolation of metabolites, DNA, long RNAs, small RNAs, and proteins from plants and microorganisms. *Plant J.* 79, 173–180. doi: 10.1111/tpj.12546
- Valledor, L., Furuhashi, T., Recuenco-muñoz, L., Wienkoop, S., and Weckwerth, W. (2014b). System-level network analysis of nitrogen starvation and recovery in *Chlamydomonas reinhardtii* reveals potential new targets for increased lipid accumulation. *Biotechnol. Biofuels* 7:171. doi: 10.1186/s13068-014-0171-1
- Valledor, L., Romero-Rodriguez, M. C., and Jorin-Novo, J. V. (2014c). Standardization of data processing and statistical analysis in comparative plant proteomics experiment. *Methods Mol. Biol.* 1072, 51–60. doi: 10.1007/978-1-62703-631-3\_5
- Wahid, A., and Close, T. J. (2007). Expression of dehydrins under heat stress and their relationship with water relations of sugarcane leaves. *Biol. Plant.* 51, 104–109. doi: 10.1007/s10535-007-0021-0
- Wahid, A., Gelani, S., Ashraf, M., and Foolad, M. R. (2007). Heat tolerance in plants: an overview. *Environ. Exp. Bot.* 61, 199–223. doi: 10.1016/j.envexpbot.2007.05.011



- Wahid, A., and Ghazanfar, A. (2006). Possible involvement of some secondary metabolites in salt tolerance of sugarcane. *J. Plant Physiol.* 163, 723–730. doi: 10.1016/j.jplph.2005.07.007
- Wang, J., Juliani, H. R., Jespersen, D., and Huang, B. (2017). Differential profiles of membrane proteins, fatty acids, and sterols associated with genetic variations in heat tolerance for a perennial grass species, hard fescue (*Festuca Trachyphylla*). *Environ. Exp. Bot.* 140, 65–75. doi: 10.1016/j.envexpbot.2017.05.014
- Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer. doi: 10.1007/978-0-387-98141-3
- Wittstock, U., and Halkier, B. A. (2000). Cytochrome P450 CYP79A2 from *Arabidopsis thaliana* L. Catalyzes the conversion of L-phenylalanine to phenylacetaldoxime in the biosynthesis of benzylglucosinolate. *J. Biol. Chem.* 275, 14659–14666. doi: 10.1074/jbc.275.19.14659
- Wu, H., Luo, D., Vignols, F., and Tsung-luo, J. (2012). Heat shock-induced biphasic  $\text{Ca}^{2+}$  signature and OsCaM1-1 nuclear localization mediate downstream signalling in acquisition of thermotolerance in rice (*Oryza sativa* L.). *Plant Cell Environ.* 35, 1543–1557. doi: 10.1111/j.1365-3040.2012.02508.x
- Xu, S., Li, J., Zhang, X., Wei, H., and Cui, L. (2006). Effects of heat acclimation pretreatment on changes of membrane lipid peroxidation, antioxidant metabolites, and ultrastructure of chloroplasts in two cool-season turfgrass species under heat stress. *Environ. Exp. Bot.* 56, 274–285. doi: 10.1016/j.envexpbot.2005.03.002
- Yoo, H., Widhalm, J. R., Qian, Y., Maeda, H., Cooper, B. R., Jannasch, A. S., et al. (2013). An alternative pathway contributes to phenylalanine biosynthesis in plants via a cytosolic tyrosine:phenylpyruvate aminotransferase. *Nat. Commun.* 4:2833. doi: 10.1038/ncomms3833
- Zhang, Q., Su, L.-J., Chen, J.-W., Zeng, X.-Q., Sun, B.-Y., and Peng, C.-L. (2012). The antioxidative role of anthocyanins in *Arabidopsis* under high-irradiance. *Biol. Plant.* 56, 97–104. doi: 10.1007/s10535-012-0022-5
- Zhang, Y., Zhao, F., Deng, Y., Zhao, Y., and Ren, H. (2015). Metagenomic and metabolomic analysis of the toxic effects of trichloroacetamide-induced gut microbiome and urine metabolome perturbations in mice. *J. Proteome Res.* 14, 1752–1761. doi: 10.1021/pr5011263
- Zulak, K. G., and Bohlmann, J. (2010). Terpenoid biosynthesis and specialized vascular cells of conifer defense. *J. Integr. Plant Biol.* 52, 86–97. doi: 10.1111/j.1744-7909.2010.00910.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 Escandón, Meijón, Valledor, Pascual, Pinto and Cañal. 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.



# Cyclotide Evolution: Insights from the Analyses of Their Precursor Sequences, Structures and Distribution in Violets (*Viola*)

Sungkyu Park<sup>1</sup>, Ki-Oug Yoo<sup>2</sup>, Thomas Marcussen<sup>3</sup>, Anders Backlund<sup>1</sup>, Erik Jacobsson<sup>1</sup>, K. Johan Rosengren<sup>4</sup>, Inseok Doo<sup>5</sup> and Ulf Göransson<sup>1\*</sup>

<sup>1</sup> Division of Pharmacognosy, Department of Medicinal Chemistry, Uppsala University, Uppsala, Sweden, <sup>2</sup> Department of Biological Sciences, Kangwon National University, Chuncheon, South Korea, <sup>3</sup> Department of Biosciences, Centre for Ecological and Evolutionary Synthesis, University of Oslo, Oslo, Norway, <sup>4</sup> School of Biomedical Sciences, The University of Queensland, Brisbane, QLD, Australia, <sup>5</sup> Biotech Research Team, Biotech Research Center of Dong-A Pharm Co Ltd., Seoul, South Korea

## OPEN ACCESS

### Edited by:

Luis Valledor,  
Universidad de Oviedo, Spain

### Reviewed by:

Jesús Pascual Vázquez,  
University of Turku, Finland  
Diego Mauricio Riaño-Pachón,  
Institute of Chemistry, University of  
São Paulo, Brazil  
Monica Escandon,  
University of Aveiro, Portugal

### \*Correspondence:

Ulf Göransson  
ulf.goransson@fkg.uu.se

### Specialty section:

This article was submitted to  
Plant Systems and Synthetic Biology,  
a section of the journal  
Frontiers in Plant Science

Received: 07 June 2017

Accepted: 17 November 2017

Published: 18 December 2017

### Citation:

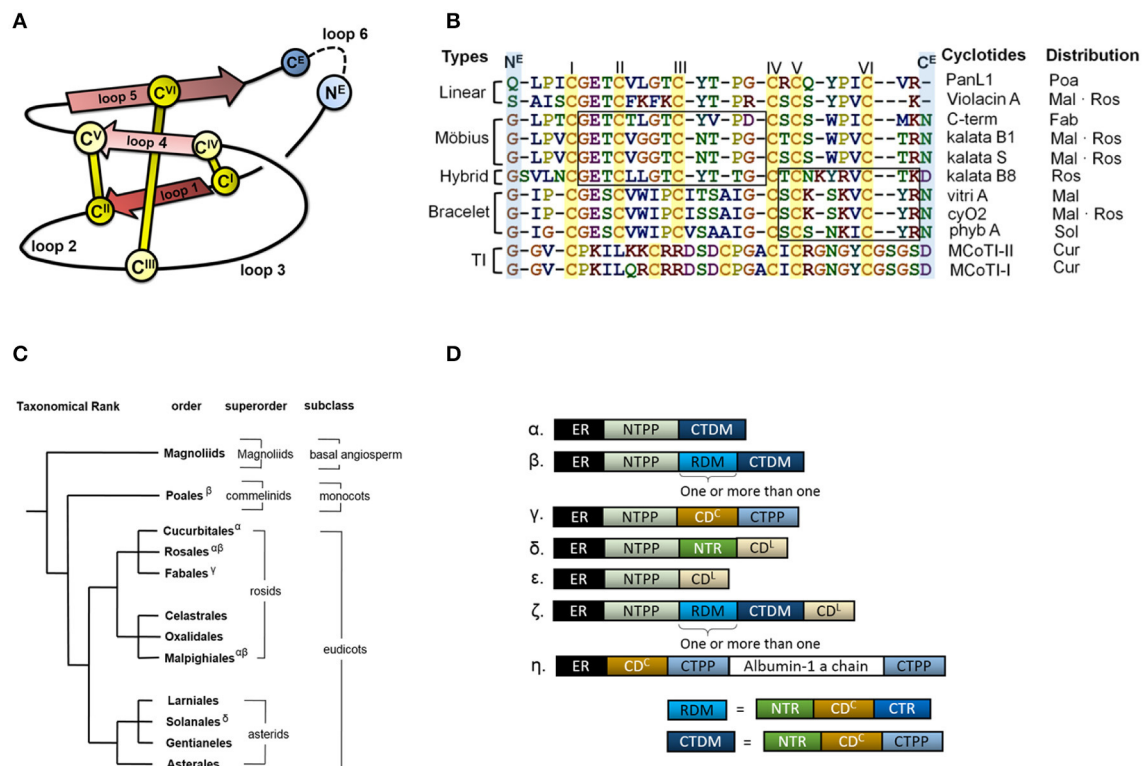
Park S, Yoo K-O, Marcussen T,  
Backlund A, Jacobsson E,  
Rosengren KJ, Doo I and  
Göransson U (2017) Cyclotide  
Evolution: Insights from the Analyses  
of Their Precursor Sequences,  
Structures and Distribution in Violets  
(*Viola*). Front. Plant Sci. 8:2058.  
doi: 10.3389/fpls.2017.02058

Cyclotides are a family of plant proteins that are characterized by a cyclic backbone and a knotted disulfide topology. Their cyclic cystine knot (CCK) motif makes them exceptionally resistant to thermal, chemical, and enzymatic degradation. By disrupting cell membranes, the cyclotides function as host defense peptides by exhibiting insecticidal, anthelmintic, antifouling, and molluscicidal activities. In this work, we provide the first insight into the evolution of this family of plant proteins by studying the Violaceae, in particular species of the genus *Viola*. We discovered 157 novel precursor sequences by the transcriptomic analysis of six *Viola* species: *V. albida* var. *takahashii*, *V. mandshurica*, *V. orientalis*, *V. verecunda*, *V. acuminata*, and *V. canadensis*. By combining these precursor sequences with the phylogenetic classification of *Viola*, we infer the distribution of cyclotides across 63% of the species in the genus (i.e., ~380 species). Using full precursor sequences from transcriptomes, we show an evolutionary link to the structural diversity of the cyclotides, and further classify the cyclotides by sequence signatures from the non-cyclotide domain. Also, transcriptomes were compared to cyclotide expression on a peptide level determined using liquid chromatography-mass spectrometry. Furthermore, the novel cyclotides discovered were associated with the emergence of new biological functions.

**Keywords:** cyclotide evolution, viola phylogeny, sequence signature, cyclotide precursor, neofunctionality, novel cyclotide, precursor domain

## INTRODUCTION

Cyclotides are proteins of ~30 amino acid residues that are characterized by the cyclic cystine knot (CCK) motif (Craik et al., 1999; Burman et al., 2014). The CCK motif consists of six conserved cysteines that form three disulfide bonds and a head to tail cyclic backbone (Figure 1A). The cyclotides have been classified into two main subfamilies, the Möbius and the bracelets, based on a single structural trait: the presence or absence of a conceptual 180° twist in the cyclic backbone caused by a conserved *cis*-Pro residue in loop 5 (Craik et al., 1999; Figure 1B). Aside from the CCK, two loops (defined as sequences between adjacent cysteines) have high sequence



**FIGURE 1 |** Cyclotide sequence, structure and distribution. **(A)** Schematic representation of the structure of a cyclotide and the CCK framework. The cyclization point between the C- and N-termini (N<sup>E</sup> and C<sup>E</sup>, respectively) is highlighted, and the cysteine residues are labeled with Roman numerals. **(B)** Sequence alignment of representative cyclotides and their distribution in flowering plants. The cyclotides have been classified into five groups on the basis of their structural traits: Möbius cyclotides, bracelets, hybrids, linear cyclotides, and trypsin inhibitors (TI). Abbreviation code in plant orders: Poaceae (Poa), Malpighiales (Mal), Rosales (Ros), Fabales (Fab), Solanales (Sol) and Cucurbitales (Cur). The loops sharing sequence similarity are boxed. **(C)** Phylogeny of cyclotide-expressing flowering plants. To date, cyclotides have only been found in the eudicots. The family Violaceae belongs to the order of Malpighiales, Rubiaceae to Gentianales, and Cucurbitaceae to Cucurbitales. **(D)** Diversity of cyclotide precursor architectures. Precursor sequences consist of multiple domains: endoplasmic reticulum targeting signal (ER, black), N-terminal propeptide (NTPP, light blue), N-terminal repeat (NTR, green), cyclotide domain (CD), and C-terminal tail (CTR). Precursor architecture varies with both plant species and structural subfamily. Architecture types are labeled with Greek letters in (D), and their distribution is shown in (C). Oak1 ( $\alpha$ ) and Vok1 ( $\beta$ ), precursors of kb1, are found in the Rubiaceae and Violaceae, respectively. Vok1 contains repeated domains. VocA ( $\delta$ ) and panitide pL1 ( $\epsilon$ ) are precursors of linear cyclotides and are found in the Violaceae and Poaceae, respectively. Precursors of these linear cyclotides do not contain a CTPP domain. TI subfamily precursors ( $\zeta$ ) contain both linear and cyclic cyclotides in the same sequence. Several architectures are often found in a single plant species. For example, *Viola* species contain Vok1 ( $\alpha$ ), Voc1 ( $\beta$ ), and VocA ( $\delta$ ). In the precursor of Cter M, the cyclotide domain replaces the albumin-1b chain.

similarity between subfamilies (loop 1 and 4), whereas loops 2 and 3 are conserved only within individual subfamilies.

The discovery of other cyclotides has created a need for a more versatile classification system (Ireland et al., 2006; Nguyen et al., 2013; Ravipati et al., 2015). These varieties include so-called hybrid cyclotides that exhibit sequence characteristics of both the Möbius and bracelet subfamilies (Daly et al., 2006), as well as minor subfamily known as the trypsin inhibitors originating from gourd plants (Hernandez et al., 2000). They contain the CCK motif but do not otherwise exhibit any sequence similarity with the other subfamilies. In addition, linear cyclotide derivatives that exhibit sequence similarity with conventional cyclotides but lack their cyclic backbone have been reported (Ireland et al., 2006; Nguyen et al., 2013). The high sequence diversity of the cyclotides appears to be due to natural selection in angiosperms, the flowering plants, but little is known about the evolutionary mechanisms underpinning the corresponding

selection processes or the evolutionary background of cyclotide diversity.

Cyclotides and the CCK motif have only been found in angiosperms, but proteins having one of their two defining structural motifs—cyclic peptides/proteins without the cystine-knot (Trabi and Craik, 2002; Arnison et al., 2013) or linear proteins with a cystine-knot (Zhu et al., 2003)—are found in a wide range of organisms across all kingdoms of life. In angiosperms, the occurrences of cyclotides differ between “basal” angiosperms, monocots and eudicots (Figure 1C): “linear cyclotides,” i.e., peptides that exhibit sequence similarity with cyclotides but lack their head-to-tail cyclic structure, are prevalent in both monocots and eudicots, but true cyclotides have been found only in eudicots (Mulvenna et al., 2006; Zhang et al., 2015a). However, neither linear nor true cyclotides have yet been found in the “basal” angiosperms. It has therefore been proposed that linear cyclotides are ancestral (more primitive) to

the true cyclic cyclotides (Mulvenna et al., 2006; Gruber et al., 2008).

To date, cyclotides have been discovered in the eudicot families of Rubiaceae (coffee) (Gran, 1973a), Violaceae (violet family) (Schöpke et al., 1993; Claeson et al., 1998), Fabaceae (legume family) (Poth et al., 2011a), Solanaceae (potato) (Poth et al., 2012), and Cucurbitaceae (cucurbit) (Hernandez et al., 2000), as well as in the monocot family Poaceae (grass family) (Nguyen et al., 2013). Cyclotides appear to have functions in host defense because they exhibit insecticidal (Jennings et al., 2001), anthelmintic (Colgrave et al., 2008a), antifouling (Göransson et al., 2004), and molluscicidal (Plan et al., 2008) activities. In addition, native cyclotides have uterotonic (Gran, 1973b), anti-neurotensin (Witherup et al., 1994), antibacterial (Tam et al., 1999; Pránting et al., 2010), anti-HIV (Gustafson et al., 1994), anticancer (Lindholm et al., 2002), and immunosuppressive (Gründemann et al., 2012) activities. This plethora of activities and the stability of the CCK motif make them of interest for drug development (Northfield et al., 2014).

Many of these activities appear to be due to cyclotides' ability to interact with and disrupt biological membranes (Colgrave et al., 2008b; Simonsen et al., 2008; Burman et al., 2011; Henriques et al., 2011). The membrane disruption is mediated by physicochemical interactions between cyclotides and the lipid membrane, and is governed by the distribution of lipophilic and electrostatic properties over the molecular surfaces of the cyclotides. We recently developed a quantitative structure-activity relationship (QSAR) model for these interactions (Park et al., 2014). However, the relationships between cyclotide sequence diversity, evolutionary selection, and the functions of the cyclotides *in planta* remain unknown.

Cyclotides are expressed as precursor proteins, which undergo post-translational processing including enzymatic cleavage and subsequent cyclization (Jennings et al., 2001; Harris et al., 2015). The multi-domain architecture of these precursor proteins varies slightly between different types of cyclotides and plant families, but in sequential order from the N- to the C-terminus, they generally feature the following domains: an endoplasmic reticulum (ER) targeting signal, an N-terminal propeptide (NTPP), an N-terminal repeat (NTR), the cyclotide domain (CD), and finally a C-terminal tail (CTR) (Figure 1D). In some cases, the modular domains NTR, CD, and CTR are repeated more than once. The cyclotides have been suggested to co-evolve with asparaginyl endopeptidase (AEP) because of its suggested role in cyclization (Mylne et al., 2012). Moreover, the divergent evolution of cyclotides from ancestral albumin domains was suggested based on the architecture of cyclotide precursors found in the Fabaceae plant family (Nguyen et al., 2011; Poth et al., 2011b). However, the relationship between the precursor proteins' architecture and sequences and the evolutionary selection of cyclotides is still unknown.

To date, cyclotide and precursor sequences have been most extensively explored in the family Violaceae Batsch. (Malpighiales), and especially in the genus *Viola* L. (Burman et al., 2015). The Violaceae are a medium-sized family including ~1,100 species worldwide. The phylogeny of the Violaceae has

recently been inferred from chloroplast and nuclear markers (Tokuoka, 2008; Wahlert et al., 2014), and its systematics has been revised accordingly; currently ~30 genera are accepted in nomenclature (Wahlert et al., 2014). *Viola* is the largest genus in the family, with 580–620 species, representing over 50% of all known species (Ballard et al., 1998; Yockteng et al., 2003; Marcussen et al., 2012; Wahlert et al., 2014). *Viola* is distributed all around the world in temperate regions and at high elevation habitats in the tropics. The genus is old (~30 million years) and comprises at least 16 extant lineages, referred to as sections, with a complex, reticulate phylogenetic history owing to allopolyploidy (Marcussen et al., 2012, 2015). The species included in this study belong to four north-temperate sections, the diploid sect. *Chamaemelanium* Ging. (*V. canadensis* L., *V. orientalis* W.Becker) and the three allotetraploid sections *Melanium* Ging. (*V. tricolor* L.), *Plagiostigma* Godr. (*V. albida* Palibin. var. *takahashii* (Nakai) Kitag., *V. mandshurica* W.Becker, *V. verecunda* A.Gray) and *Viola* (*V. acuminata* Ledeb.).

In the current study, we explore cyclotide evolution using an integrated approach, exploiting transcriptomics and peptidomics to analyze the sequences of cyclotide precursors and the expression of cyclotides in *Viola* in light of the phylogeny of the genus. In particular, full precursor sequences are used to obtain insights into the evolutionary history of the cyclotides, and we connect the evolution of new mature cyclotides with the emergence of new functions.

## MATERIALS AND METHODS

### Collection of Violets

Violets were collected at the sites indicated in Table 1, and those sites are their natural habitats. When collected, the plant individuals were in adult stage; however, the exact ages of those plant individuals were not determined. The plant vouchers were deposited at the Kangwon National University herbarium; *Viola albida* var. *takahashii* (KWNU93021), *V. mandshurica* (KWNU93022), *V. orientalis* (KWNU93023), *V. verecunda* (KWNU93024), and *V. acuminata* (KWNU93025). All plant material was collected on September 23, 2014.

### Sample Collection, RNA Isolation, and RNA Sequencing

For the five *Viola* species—*V. albida* var. *takahashii*, *V. mandshurica*, *V. orientalis*, *V. verecunda*, and *V. acuminata*—, total RNA was sequenced by Next Generation Sequencing (NGS), outsourced to Macrogen Inc. (Seoul, South Korea). For each *Viola* species, the RNA sample was prepared from one plant individual; also, the plant tissues were pooled from all of the plant's major organs, i.e., the roots, stems, flowers, and leaves. The collected tissues were immediately frozen in liquid nitrogen, and directly extracted by RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to the manufacturers' protocols. Quality and quantity of RNA were measured using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) with an RNA Integrity Number (RIN) and rRNA ratio. The measured RIN values (rRNA ratios) are: 6.9 (1.2)



**TABLE 1** | The origins of the plant material examined in this work.

Plants collected (Scientific name)	Locality	Geographic Coordinates and altitudes
<i>V. albida</i> var. <i>takahashii</i>	Mt. Cheonma, Namyangju-si, Gyeonggi-do, South Korea	N37°41'43.0" E127°17'25.5", 213 m
<i>V. mandshurica</i>	Kangwon National University Campus, Chuncheon-si, Gangwon-do, South Korea	N37°52'25.9" E127°44'68.6" 110 m
<i>V. orientalis</i>	Mt. Daeryong, Sumit area, Chuncheon-si, Gangwon-do, South Korea	N37°50'68.8" E127°49'12.5" 860 m
<i>V. verecunda</i>	Mt. Daeryong, Valley area, Chuncheon-si, Gangwon-do, South Korea	N37°50'03.5" E127°49'80.7" 415 m
<i>V. acuminata</i>	Mt. Oeum, Hoengsung-gun, Gangwon-do, South Korea	N37°36'34.9" E127°56'12.9" 619 m

for *V. albida* var. *takahashii*, 7.4 (1.6) for *V. mandshurica*, 7.2 (0.2) for *V. orientalis*, 6.4 (0.9) for *V. verecunda* and 5.9 (0.7) for *V. acuminata*. For mRNA library preparation, a TruSeq RNA preparation kit was used according to the manufacturer's instructions (Illumina, San Diego, U.S.A.). Briefly, the poly-A containing mRNAs were purified using poly-T oligo-attached magnetic beads. The purified mRNAs were fragmented into short sequences by use of divalent cations. Using these short sequences as templates, the first-stranded cDNA was synthesized using random hexamers. A second-strand cDNA was then synthesized using DNA polymerase I and RNase H. The synthesized cDNA went through an end repair process, the addition of a single "A" base, and then ligation of the adapters. The PCR was performed to enrich the selected DNA sequences, and then those selected sequences were sequenced using Illumina HiSeq 2000 Sequencing System (Illumina, USA) that generates paired-end reads with  $2 \times 100$  base pairs (bp) read length.

## Sequencing Data Analysis and Assembly

FASTQC (version 0.11.3) was used to determine the quality of RNA sequencing data ([www.bioinformatics.babraham.ac.uk/projects/fastqc/](http://www.bioinformatics.babraham.ac.uk/projects/fastqc/)). Reads were cleaned using Trimmomatic (v. 0.32; Bolger et al., 2014), and the sequences with Phred score  $\geq 33$  and a minimum length of 36 bp were retained for assembly. *De novo* assembly of these processed reads was performed using the Trinity RNA-Seq assembly (release 17.07.2014; <https://sourceforge.net/projects/trinityrnaseq/>) with the default setup, which allows the identification of cyclotide-coding transcripts larger than 200 bp. Also, the reads were assembled separately for the individual species. We summarized the detailed information on the *de novo* transcriptome assembly for those individual species in Supplementary Table 1. The transcript abundance, Fragments Per Kilobase of transcript per Million mapped reads (FPKM; Trapnell et al., 2010), was calculated using Trinity based on RSEM algorithm (Li and Dewey, 2011).

## Identification of Precursor Sequences from Transcriptome

The assembled transcriptomes were searched for similar sequences to the cyclotides from Cybase (cutoff date: 3.02.2014; <http://www.cybase.org.au/>) by the standalone NCBI-blast+ service (2.2.28) (tblastn, E-value cutoff: 50) in the Ugene software package (Okonechnikov et al., 2012). In parallel, a

motif search was done using PROSITE for cyclotide precursor sequences [motifs PS60009 and PS60008 for Möbius and bracelet cyclotides, respectively (Sigrist et al., 2002)]. To assist sequence identification, the Fuzzpro service of EMBOSS (v. 5.0.0) (Rice et al., 2000) was used. The combined PROSITE and blast results were filtered to remove duplicates. The result file was further processed by manual inspection after clustal omega (v.1.2.0) alignment. In the manual inspection, we assumed that the sequence is a cyclotide precursor if the sequence contains the six conserved cysteines aligned with previously known cyclotides, and if the conserved N-terminal domain showed sequence similarity to known cyclotide precursors.

## Collection of Precursor Sequences

In total, 312 (= 157 + 155) precursor sequences were utilized in this study: (i) 157 (=138 + 19) precursor sequences were identified and collected from transcriptome analyses conducted in the current study. Among them, 138 sequences were collected from the five *Viola* species (i.e., *Viola albida* var. *takahashii*, *V. mandshurica*, *V. orientalis*, *V. verecunda*, and *V. acuminata*) by the RNA sequencing in this study. The remaining 19 sequences were collected the transcriptomic data of *V. canadensis* obtained from the 1kp-project ([www.onekp.com](http://www.onekp.com)). (ii) Another 155 (=126 + 29) precursor sequences were collected from recent published studies. Among them, the 126 sequences were collected from other seven *Viola* species (i.e., *V. baoshanensis*, *V. odorata*, *V. uliginosa*, *V. adunca*, *V. tricolor*, *V. biflora*, and *V. pinetorum*), and the 29 sequences from other Violaceae genera, i.e., *Gloeospermum*, *Melicytus*, and *Pigea*. The name of *Viola* species where the precursor sequences collected from and the corresponding references are listed in the Table 2, and the sequence alignment of those 312 precursors are recorded in the Supplementary Data 1.

## Nomenclature of Cyclotides and Precursors

Each cyclotide sequence was assigned a tripartite name. The first part is derived from the Latin binomial of the plant in which the corresponding precursor was found, the third part specifies the molecular species of the cyclotide, and the second part specifies the rank of the cyclotide among all the cyclotides having that particular molecular species that were identified in that particular

**TABLE 2 |** Violaceae species sampled for the analysis of precursor sequences.

Genus	Section	Species
<i>Viola</i>	<i>Plagiostigma</i>	<b><i>V. mandshurica</i> W.Becker</b> [37] <sup>a</sup> , <b><i>V. albida</i> Palib. var. <i>takahashii</i> (Nakai) Kitag.</b> [29] <sup>a</sup> , <b><i>V. verecunda</i> A.Gray</b> [23] <sup>a</sup> , <i>V. baoshanensis</i> W.S.Shu, W.Liu & C.Y.Lan [44] <sup>b</sup>
	<i>Viola</i>	<i>V. odorata</i> L. [5] <sup>c</sup> , <i>V. uliginosa</i> Bess. [18] <sup>d</sup> <b><i>V. acuminata</i> Ledeb.</b> [23] <sup>a</sup> <i>V. adunca</i> Sm. [2] <sup>e</sup>
	<i>Melanium</i>	<i>V. tricolor</i> L. [49] <sup>f</sup>
	<i>Chamaemelianum</i>	<b><i>V. orientalis</i> W.Becker</b> [26] <sup>a</sup> , <i>V. biflora</i> L. [6] <sup>g</sup> , <b><i>V. canadensis</i> L.</b> [19] <sup>h</sup> , <i>V. pinetorum</i> Greene [2] <sup>e</sup>
<i>Gloeospermum</i>		<i>G. blakeanum</i> (Standl.) Hekking. [5] <sup>i</sup> , <i>G. pauciflorum</i> Hekking. [4] <sup>j</sup>
<i>Melicytus</i>		<i>M. ramiflorus</i> J.R.Forst. & G.Forst. [14] <sup>j</sup>
<i>Pigea</i>		<i>P. floribunda</i> Lindl. (= <i>Hybanthus floribundus</i> (Lindl.) F.Muell.) [6] <sup>k</sup>

The numbers in parentheses after the species name show the number of precursor sequences derived from that species. Species names marked in bold were used for transcriptomic assay of the current study. Sections are indicated within *Viola*. Within sect. *Plagiostigma*, *V. mandshurica*, and *V. albida* var. *takahashii* belong in subsect. *Patellares* (Boiss.) Rouy & Foucaud, *V. verecunda* in subsect. *Bilobatae* (W.Becker) W.Becker and *V. baoshanensis* in subsect. *Diffusae* (W.Becker) Chang. Within sect. *Viola*, *V. odorata* belongs to subsect. *Viola* and *V. acuminata*, *V. adunca*, and *V. uliginosa* in subsect. *Rostratae* (W.Becker) W.Becker. No subsections are applicable to sect. *Chamaemelianum*. <sup>a</sup>Transcriptomes from the current study; *V. mandshurica*, *V. albida* var. *takahashii*, *V. verecunda*, *V. acuminata* and *V. orientalis*; <sup>b</sup>*V. baoshanensis* (Zhang et al., 2009, 2015b); <sup>c</sup>*V. odorata* (Dutton et al., 2004; Ireland et al., 2006); <sup>d</sup>*V. uliginosa* (Slazak et al., 2015); <sup>e</sup>*V. pinetorum* and *V. adunca* (Kaas and Craik, 2010); <sup>f</sup>*V. tricolor* (Mulvenna et al., 2005; Hellinger et al., 2015); <sup>g</sup>*V. biflora* (Herrmann et al., 2008); <sup>h</sup>Transcriptome data for *V. canadensis* were obtained from the 1kp project ([www.onekp.com](http://www.onekp.com)), and the cyclotide precursor sequences were determined in this work; <sup>i</sup>*G. blakeanum* and *G. pauciflorum* (Burman et al., 2010); <sup>j</sup>*M. ramiflorus* (Trabi et al., 2009); <sup>k</sup>*P. floribunda* (as *H. floribundus*) (Simonsen et al., 2005).

plant species (Supplementary Table 2). The Latin binomials of the violet species considered in this work are *Viola albida* var. *takahashii* (valta), *Viola mandshurica* (viman), *Viola orientalis* (vorie), *Viola verecunda* (viver), and *Viola acuminata* (vacum). Thus, the cyclotide named *vacum2-HS4* is the second cyclotide of the HS4 molecular species derived from *V. acuminata*. The molecular species are named after their NTR signature sequences, i.e., the identities of the two residues at the consensus positions 11 and 12. The precursor sequences were also assigned tripartite names in a similar way to the cyclotides: the first part of the name is derived from the Latin binomial of the species in which the sequence was discovered, the second specifies the numerical rank of the sequence (which is independent of that assigned to the cyclotides), and the third specifies the molecular species of the precursor. Any cyclotide or precursor that had been named in an earlier publication was assigned the same name in this work. However, for previously unnamed precursors, we added the prefix “prc” to the tripartite name to help readers distinguish between precursor and cyclotide sequences. Thus, prc-viul A is the name of the precursor sequence of Viul A from *Viola uliginosa* Bess.

## Sequence Alignment for Phylogenetic Analysis

The cyclotide precursor sequences from transcriptome were aligned before the phylogenetic analysis i.e., construction of Bayesian phylogenetic tree, maximum parsimony and splits networks. A total of 92 precursor sequences (two sequences from each of the 46 molecular species) were prepared as DNA sequence for the sequence alignment, and the sequences include only three domains of the precursor, i.e., NTPP, NTR, and cyclotide domains. To guide the alignment of DNA sequences of the precursors, they were translated into protein sequences and aligned independently for each molecular species using Clustal in BioEdit v.7.2.5 (Hall, 1999), and those alignments were in turn combined and realigned (i.e., keeping the indel positions

from the alignment of each of the molecular species). The resulting protein-guided alignment of nucleotide sequences was then subject to manual adjustments within reading frames. The aligned DNA sequences are shown in the Supplementary Data 2.

## The Construction of Bayesian Phylogenetic Tree

Nucleotide substitution model was selected based on the AICc criterion using JModelTest v.2.1.10 (Guindon and Gascuel, 2003; Darriba et al., 2012). A Bayesian phylogenetic analysis was conducted in BEAST v.1.7.4 (Drummond and Rambaut, 2007; Drummond et al., 2012). The analysis file was set up in BEAUti (part of the BEAST package) with the following priors: GTR+G as substitution model with empirical frequencies and four gamma categories, a lognormal relaxed clock prior with rate set to 1.0, and a Yule tree prior (birth-only process). Two MCMC chains were run for 50 million generations each, using a BEAGLE library, and parameters logged every 10,000 generations. We checked the two chains for proper mixing and convergence (i.e., ESS >500) in Tracer v.1.6.0 (<http://tree.bio.ed.ac.uk/software/tracer/>), removed a visually determined burn-in of 1 million generations from each and merged the two chains in LogCombiner v.1.7.4 (part of the BEAST package), and summarized the data in a maximum credibility tree with mean node heights using TreeAnnotator v.1.7.4 (part of the BEAST package). The resulting tree was visualized and edited in FigTree v.1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>). The input file for the BEAST and the resulting phylogenetic tree are found in the Supplementary Data 3, 4, respectively.

## The Construction of Splits Networks

In order to visualize the data, neighbor splits-networks of uncorrected P distance were produced, both for the nucleotide alignment and for the translated alignment, using SplitsTree (Huson and Bryant, 2006).

## Extraction of Plant Material

Between 250 and 500 mg of dried plant material was homogenized and incubated overnight in 6 ml of 60% acetonitrile in water containing 0.05% trifluoroacetic acid (TFA) to extract cyclotides. Extracts were lyophilized and then dissolved in 2.5 ml of solvent A (Milli-Q H<sub>2</sub>O with 0.05% TFA). Redissolved extracts were then subjected to gel filtration using PD-10 columns (GE Healthcare) according to the manufacturer's instructions to remove small molecules. The high molecular weight fractions were collected, lyophilized and then dissolved in solvent A to a concentration proportional to the original amount of extracted material (2 µl/1 mg) for LC-MS analysis.

## LC-MS

The samples were analyzed using ultra performance liquid chromatography coupled to quadrupole time-of-flight mass spectrometry (nanoAcquity UPLC/QToF Micro; Waters, Milford, MA). Samples were eluted using a gradient of acetonitrile (1 to 90% over 50 min) containing 0.1% formic acid. A nanoLC column (Waters BEH, 75 µm (i.d) × 150 nm) operated at 0.3 µl/min flow rate was used. The capillary temperature was set at 220°C and the spray voltage at 4 kV. The mass-to-charge (*m/z*) range was set from 1,000 to 2,000. LC-MS chromatogram and MS spectra were analyzed with the help of MassLynx V4.1 (Waters, Milford, MA).

## Reduction and Alkylation

Dried peptide extracts were reduced in a buffer containing 0.05 M Tris-HCl, pH 8.3, 4.2 M guanidine-HCl, and 8 mM DTT. The extraction solutions were incubated at 37°C for 2 h in the dark, followed after O<sub>2</sub> removal with Nitrogen gas. The reduced peptides were then further alkylated in a buffer containing 0.2 M Tris-HCl, pH 8.3, and 200 mM iodoacetamide for 1 h.

## Identification of Cyclotides from the Plant Extracts

On the original LC-MS chromatogram, all chromatographic peaks were manually investigated if they include a cyclotide-like mass spectrum. We assumed that peaks stemmed from cyclotide-like substances if their masses fell within the range of 2,700–3,300 Da, as deconvoluted from their doubly- and triply charged ions. Then, the presence of three disulfide bonds was used to support their identification as cyclotides: if peaks in the extract showed an increase in mass by 348.18 Da, we considered that the peak stemmed from a true cyclotide (Figure 2).

## Matching Transcriptomic Data to Cyclotides on the Protein-Level

Presence of possible cyclotides identified in precursor sequences from transcriptome analyses was determined as follows: putative cyclotide sequences were listed and their monoisotopic mass were calculated. The listed cyclotide sequences were comprised of N-terminal residues ranging from [−3, 2] on the consensus position of their precursor sequences. The C-terminal residues were found differently between cyclic cyclotides and linear cyclotides, i.e., the highly conserved N/D residue located at loop 6 for cyclic cyclotides, and the residue located next to

stop codon in their mRNA sequences for linear cyclotides. These calculated monoisotopic mass were then compared to the observed monoisotopic mass from the identified cyclotides in plant extracts using LC-MS. We regarded that the transcriptomic cyclotides were expressed in the protein-level if the difference of the monoisotopic mass <0.40 Da.

## Determination of Cyclotides' Abundance Levels in the Proteome

For each cyclotide-like substance, the chromatographic peaks were identified together with their own retention time from the original LC-MS chromatogram. These chromatographic peaks were further investigated in relation with their mass spectral peaks to estimate the signal intensity (SI) of the cyclotide-like substance. The SI is estimated as a summed signal intensity of triple-charged mass spectral peaks. According to the SI, we assigned the cyclotide abundance into three levels, i.e., the abundance level is: low if SI <250, high if SI > 1,000, and medium if 250 < SI < 1,000.

## The Calculation of Molecular Descriptors

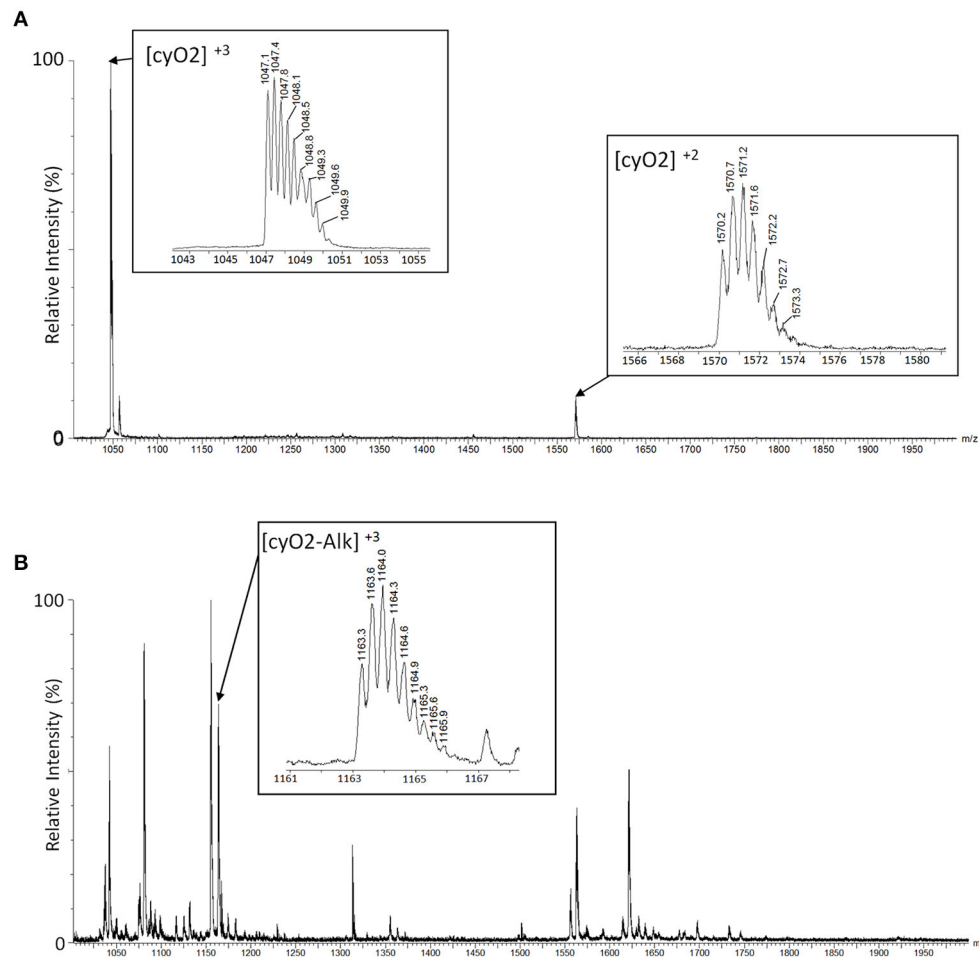
The physicochemical properties of selected cyclotides (Park et al., 2014), i.e., the total lipophilicity and the exposure ratio, were calculated using scientific vector language (SVL), implemented in MOE 2012 (Chemical Computing Group Inc., Montreal, Canada).

## Accession Numbers

The accession numbers of Sequence Read Archive (SRA) and Transcriptome Shotgun Assembly (TSA) for transcriptome of the five *Viola* species are: *V. acuminata* (TSA: GFWD000000000, SRA: SRR5320546), *V. orientalis* (TSA: GFXR000000000, SRA: SRR5322130), *V. albida* var. *takahashii* (TSA: GFWC000000000, SRA: SRR5320531), *V. mandshurica* (TSA: GFWG000000000, SRA: SRR5320533), and *V. verecunda* (TSA: GFWF000000000, SRA: SRR5322180).

## RESULT AND DISCUSSION

The family Violaceae and the genus *Viola* are an excellent system for studying the evolution of cyclotides. All investigated species express large numbers of these proteins (Burman et al., 2015) as well as their precursor sequences. Furthermore, the species-level phylogeny of *Viola* (Marcussen et al., 2012, 2015) and chloroplast phylogeny of Violaceae (Wahlert et al., 2014) are well-understood as a result of recent studies. In the current study, 312 (= 157 + 155) sequences were analyzed in total to investigate the evolution of cyclotides. Among them, 157 precursor sequences were discovered from the transcriptomes of the six *Viola* species, i.e., *V. albida* var. *takahashii*, *V. mandshurica*, *V. orientalis*, *V. verecunda*, *V. acuminata*, and *V. canadensis* (listed in the Supplementary Table 3). Those discovered 157 sequences were pooled with the other 155 precursor sequences recently found from the other *Viola* species (i.e., *V. baoshanensis*, *V. odorata*, *V. uliginosa*, *V. adunca*, *V. tricolor*, *V. biflora*, and *V. pinetorum*) and from other Violaceae genera (i.e., *Gloeospermum*, *Melicytus*, and *Pigea*).



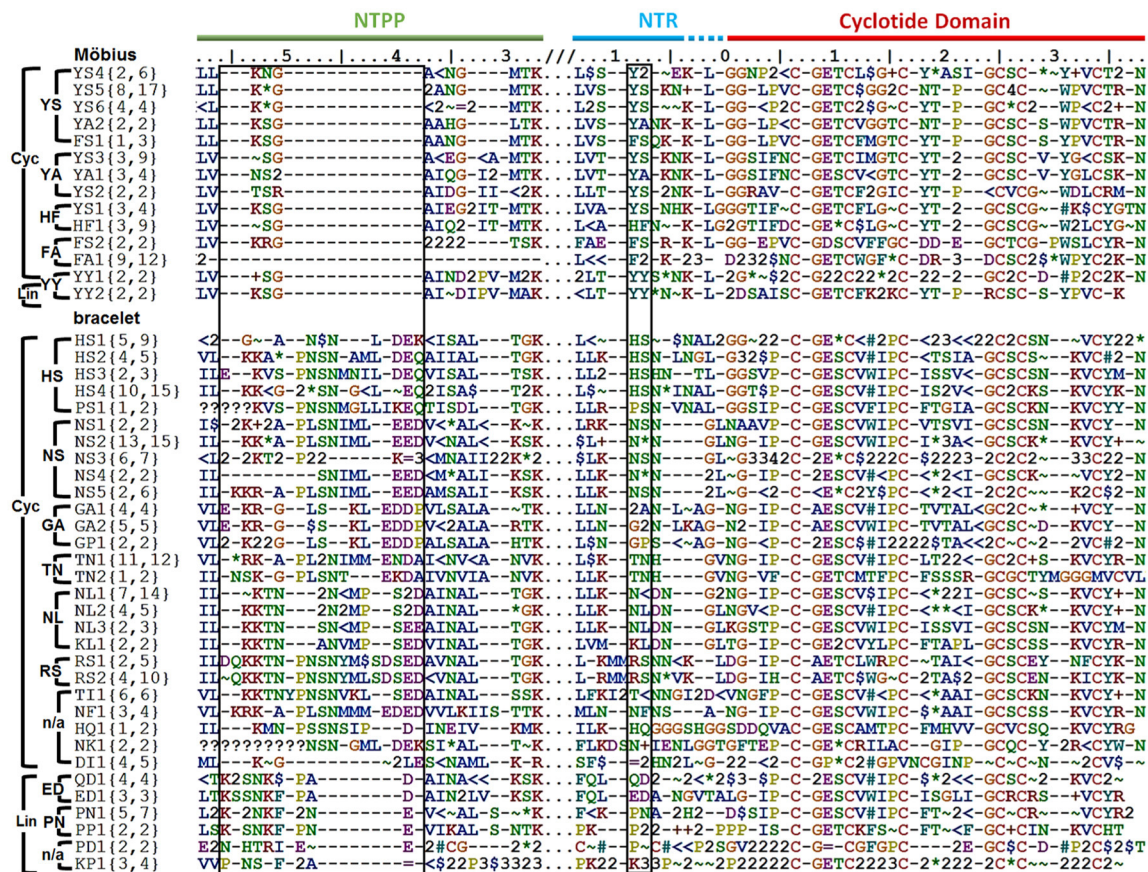
**FIGURE 2 |** Identification of cyclotides by alkylated mass changes on LC-MS. Mass spectra of a cyclotide (cyO2) are shown with a comparison between the mass spectra of the plant extract before alkylation **(A)** and after alkylation **(B)**. **(A)** We identified the mass spectra of the cyclotide-like using the mass distance on their isotopic peaks. The triple-charged monoisotopic mass of the cyO2 (marked as [cyO2]<sup>+3</sup> in the box) is observed as 1047.1Da, and the double-charged monoisotopic mass ([cyO2]<sup>+2</sup>) is observed as 1570.2Da. In each mass spectrum, the distances between isotopic peaks are observed with ~0.3Da for [cyO2]<sup>+3</sup>, and 0.5Da for [cyO2]<sup>+2</sup>. The monoisotopic mass of the cyO2 is 3138.4Da (calculated) and 3138.3Da (observed), and the mass difference is 0.1Da. **(B)** The monoisotopic mass of the cyO2 after the alkylation reaction is 3486.6Da (calculated) and 3486.9Da (observed), and the mass difference is 0.3Da. The calculated mass is based on the triple-charged monoisotopic mass of cyO2 (1163.3Da), and its related mass spectrum is shown in the box marked [cyO2-Alk]<sup>+3</sup>.

## The Sequence Signature of the Prodomain Can Be Used for the Classification of Cyclotides

The structural classification of cyclotides into Möbius cyclotides, bracelets, and hybrids thereof based on the structure and sequence of the mature cyclotide is here replaced by a classification based on the sequences of the cyclotide precursors, including both the prodomain (i.e., the NTPP and NTR domains) and the cyclotide domain. We focused on these three domains (i.e., the NTPP, NTR, and cyclotide domains) because they are present in all of the precursors that have been completely sequenced. The ER domain was not included in the analysis because ER domain sequences often vary strongly with the quality of the sequencing data. Other domains (i.e., the CTR domain and repeats of the NTR and cyclotide domains) are not present in all precursors, and were therefore also excluded.

Precursors were classified by their *sequence signatures*, i.e., the patterns of insertions and deletions (indels) and conserved sequences in the prodomains (NTPP and NTR). The N-terminal cleavage site of the cyclotide domain was defined as position 0, and a large indel region was detected upstream in the NTPP at positions [−56, −38] (**Figure 3**). Within this indel, the insertion region features sequence variations with minor gaps, and the deletion region has definite sequence gaps [−56, −54], [−50, −38]. Interestingly, the insertions coincide with cyclotide domains of archetypical bracelet cyclotides (e.g., cycloviolacin O2), while deletions coincide with cyclotide domains of archetypical Möbius cyclotides (e.g., kalata B1). Based on these observations, we suggest that these indels in the NTPP domain can be used as a criterion for classifying precursors into the Möbius and bracelet *lineages*. The term lineage is used instead of subfamily in order to avoid confusion with the structural classification of the cyclotides.





**FIGURE 3 |** Classification of cyclotide precursors based on the sequence signature. Consensus sequences of the molecular species identified in this work. Based on the sequence signatures in the prodomain, 249 cyclotide precursors were classified into the Möbius and bracelet lineages, 13 molecular series and 46 molecular species (14 Möbius and 32 bracelet species). Species are named for their signature sequences, and numbers within brackets show the numbers of complete and total (complete and partial) precursor sequences for each species. The left box shows the NTPP indel region [−56, −38], which contains defining sequence gaps at [−56, −54], and [−50, −38] in the Möbius lineage. Within a lineage, specific prodomain sequence-traits are associated with different structural subfamilies. In the Möbius lineage, the sequence (Y<sup>+</sup>/F<sup>+</sup>/H<sup>+</sup>)<sup>−9</sup>-(S<sup>+</sup>/A/Y)<sup>−8</sup> at position [−9, −8] is associated with cyclic cyclotide precursors (where “<sup>+</sup>” indicates high occurrence of the labeled residue), while Y<sup>−9</sup>-Y<sup>−8</sup> is associated with linear cyclotide precursors. Precursors of structural hybrids exhibit have an insertion at [−32, −31], whereas the archetypical Möbius precursors have a deletion at this position. In the bracelet lineage, the sequences (H<sup>+</sup>/N<sup>+</sup>/S/T/G/K/P)<sup>−9</sup>-(L/N/S/F/A)<sup>−8</sup> and (Q/E/P/K)<sup>−9</sup>-(D/N)<sup>−8</sup> are associated with cyclic and linear precursors, respectively. In addition, linear precursors contain a deletion that is flanked by two defining insertions, (P<sup>+</sup>/A/L)<sup>−49</sup>-(N<sup>+</sup>/A)<sup>−48</sup> and (D/E)<sup>−39</sup>, in the NTPP at [−49, −48] and [−39]. Conserved amino acids are indicated by their one-letter codes, and variable residues are represented by symbols indicating the physicochemical properties of their side chains: hydrophilic (~), hydrophobic (\$), flexible (G), rigid (P), disulfide-forming (C) and indels (−). Hydrophilic residues are further divided into positively charged (+), negatively charged (=) and uncharged hydrophilic (\*); hydrophobics (\$) are divided into aromatic (#) and alkyl (<) groups. Numbers are used as consensus symbols for multiple physicochemical categories.

Within each lineage, precursors can be further divided by analogy to their structural classifications. That is to say, it is possible to identify sequence traits associated with linear and cyclic cyclotides within each lineage, and sequence traits associated with archetypical and hybrid subfamilies within the cyclic cyclotides of the Möbius lineage. Signature sequences of residues [−56, −38] were used to subdivide the bracelets, and residues [−32, −31] for the Möbius lineages, together with residues [−9, −8] of the NTR. Within the bracelet lineage, the sequence signatures associated with linear and cyclic cyclotide precursors differ in two ways. First, linear cyclotide precursors contain a signature sequence in the NTPP domain defined by two sequence insertions at [−49, −48] and [−39], together with

a sequence deletion between those two sequence insertions. The signature sequences at these two locations are (P<sup>+</sup>/A/L)<sup>−49</sup>-(N<sup>+</sup>/A)<sup>−48</sup> and (D/E)<sup>−39</sup>, where the symbol “<sup>+</sup>” denotes a residue that occurs most frequently at the indicated position. Conversely, precursors of cyclic peptides have variable sequences with minor gaps in the [−47, −40] region. Second, whereas positions [−9, −8] in the NTR of precursors of linear cyclotides are conserved, having the signature sequence (Q/E/P/K)<sup>−9</sup>-(D/N)<sup>−8</sup>, the precursors of cyclic cyclotides show higher diversity, having sequences of (H<sup>+</sup>/N<sup>+</sup>/S/T/G/K/P)<sup>−9</sup>-(L/N/S/F/A)<sup>−8</sup>. Different signature sequences were also identified between the three structural subfamilies of the Möbius lineage, i.e., linear, hybrid and archetypical Möbius cyclotides. First, while precursors of

both linear and hybrid cyclotides contain a sequence insertion in the NTPP domain at [−32, −31], precursors of archetypical Möbius cyclotides contain a deletion at this position. Second, while the sequence at [−9, −8] in the NTR domain of linear cyclotides is highly conserved ( $Y^{-9}-Y^{-8}$ ), it is more variable in the precursors of cyclic cyclotides ( $Y^{\wedge}/F^{\wedge}/H)^{-9}-(S^{\wedge}/A/Y)^{-8}$ .

On the basis of these findings, i.e., the combined sequence signatures of the NTPP and NTR domains and the conservation of residues with similar physicochemical properties in the cyclotide domain, two new classification orders were defined and used to classify the members of each lineage. These new orders were termed the molecular species and molecular series. Precursors with the same sequence signatures in both the NTPP and NTR domains are assigned to the same molecular species, while precursors that only share signature sequences in the NTR domain are assigned to the same molecular series.

Of 283 precursor sequences, 80 were classified into the Möbius lineage and 181 into the bracelet lineage (Figure 3; see Supplementary Figure 1). Of these 261 sequences, 249 were classified into 13 molecular series and 46 molecular species. Thus, 78 sequences representing the Möbius lineage were classified into five molecular series and 14 molecular species, while 171 sequences representing the bracelet lineage were classified into 8 molecular series and 32 molecular species. The remaining precursor sequences (34 of the initially examined 283, or 12.0% of the total) could not be grouped with any other sequences, and not be classified with molecular species (Supplementary Figure 2). Most of these sequences (31/283, or 10.9% of the total) were either partial sequences or lone unique sequences. Only three of all the precursor sequences (3/283) did not exhibit features enabling their classification into a given lineage on the basis of their NTPP domain sequences.

Informal hierarchical ranks were then assigned to this system for classifying precursors, with lineage being the highest ranking classification, followed by molecular series and then molecular species. We further investigated the evolutionary relevance of this classification system by performing the phylogenetic analysis using the full DNA precursor sequences (Figure 4). We assumed that if the sequence signatures are highly conserved in the course of the evolution, the phylogenetic relationship of the precursor sequences has the hierarchical ranks as classified based on the sequence signatures. In this evaluation, we adopted a Bayesian phylogenetic approach using BEAST, because substitution model-based methods are in most cases better than Maximum Parsimony or distance-based methods at handling homoplasy, which is prominent among the cyclotide precursor sequences (Splits network shown in Supplementary Figure 3), and because BEAST sets a prior on the branch lengths that is particularly suitable also for analysis of short DNA sequences, such as cyclotide precursors.

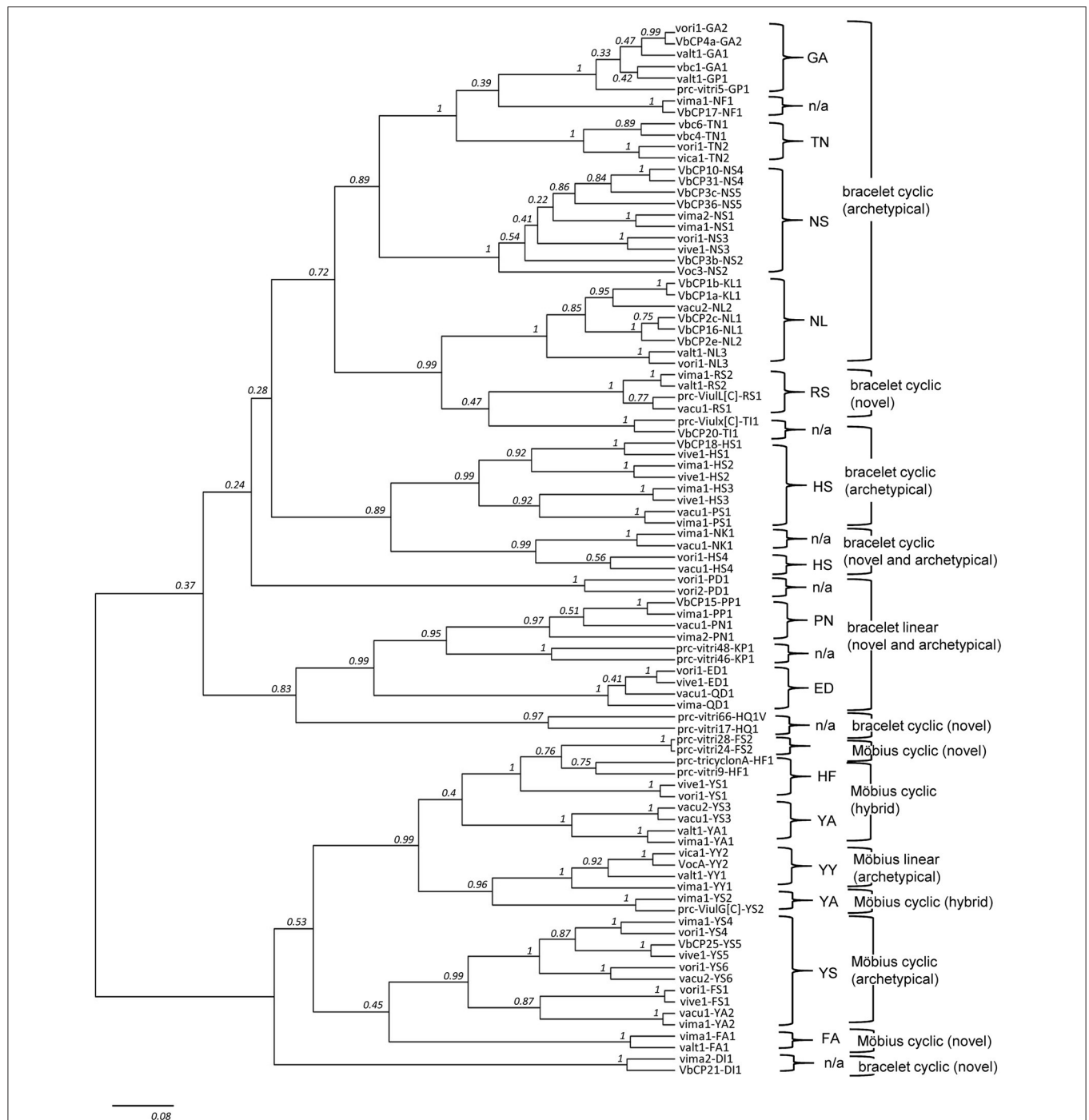
The two lower hierarchical ranks (i.e., molecular species and series) are consistently recovered in the phylogeny, but not the upper hierarchical rank (i.e., molecular lineage), as indicated by the low posterior support values at the base of the phylogeny (Figure 4). Most of precursor sequences belong to the same molecular species are grouped as one monophyletic clade (80%, 74/92) or to the different clades but more closely than the

precursor sequences derived from the different molecular series. Among 13 molecular series, the molecular species belong to the four molecular series (i.e., HF, YA, FA, and HS) are not monophyletic. Only one molecular species (DI1) is exceptionally grouped into Möbius lineage in the phylogenetic tree, even though the DI1 is classified as the bracelet lineage based on the sequence signature. In the phylogenetic analysis, we selected a total of 92 precursor sequences by the random selection of two sequences from each of the 46 molecular species. We assumed that such precursor selection would be enough to show a phylogenetic support for the classification system based on the sequence signature approach, because we randomly selected two sequences from each molecular species, and these sequences were mostly paired or grouped into monophyletic clades in accordance with the classification at the level of molecular species and series.

## The Distribution of Cyclotide Precursors Reflects the Phylogeny of the Genus *Viola*

Precursor sequences were compared to the established phylogenetic relationships between the four infrageneric sections of *Viola* (i.e., sects. *Melanium*, *Plagiostigma*, *Viola*, and *Chamaemelanium*), and between the genus *Viola* and the other *Violaceae* genera. The presence of a molecular species in different *Viola* taxa, such as species or sections, indicates that the molecular species evolved prior to the most recent common ancestor of these taxa. However, such inferences in *Viola* are complicated by the network-like phylogeny of the genus, owing to repeated ancient events of allopolyploidy (Marcussen et al., 2015). Hence, out of four sections in the current analysis, three (i.e., sects. *Melanium*, *Plagiostigma*, *Viola*) are allotetraploids originated by independent hybridizations between the same two parental lineages MELVIO and CHAM around 15 million years ago. The last section, sect. *Chamaemelanium*, is diploid and descends from the CHAM lineage. Because the parental MELVIO lineage is now extinct, it is impossible to infer with certainty the distribution of the molecular species in the common ancestor of these parental lineages, i.e., the CHAM and MELVIO lineages, especially without diploid outgroups from sister sections (e.g., *Rubellium* and *Andinium*). Gene flow by introgression between species potentially occurs between closely related species only, i.e., within the same subsection, and can be ruled out for this dataset (e.g., Marcussen et al., 2015).

The current analysis revealed several points about the distribution of cyclotide precursors within the genus *Viola*. Firstly, the classification into infrageneric sections is reflected in the distribution of the cyclotide precursors. Some molecular species occurred sporadically or commonly across the four infrageneric sections: 2% of molecular species were found across all sections, and 5% of the studied precursor sequences belong to those molecular species (Figure 5; see Supplementary Figure 4). Also, 17% of molecular species (34% of precursor sequences) were found across three sections, and 45% of molecular species (41% of precursor sequences) were found in at least two sections. This indicates that the molecular species likely originated both from the common ancestor of *Viola* sections and from the hybridization between the parental lineages. Also, there are

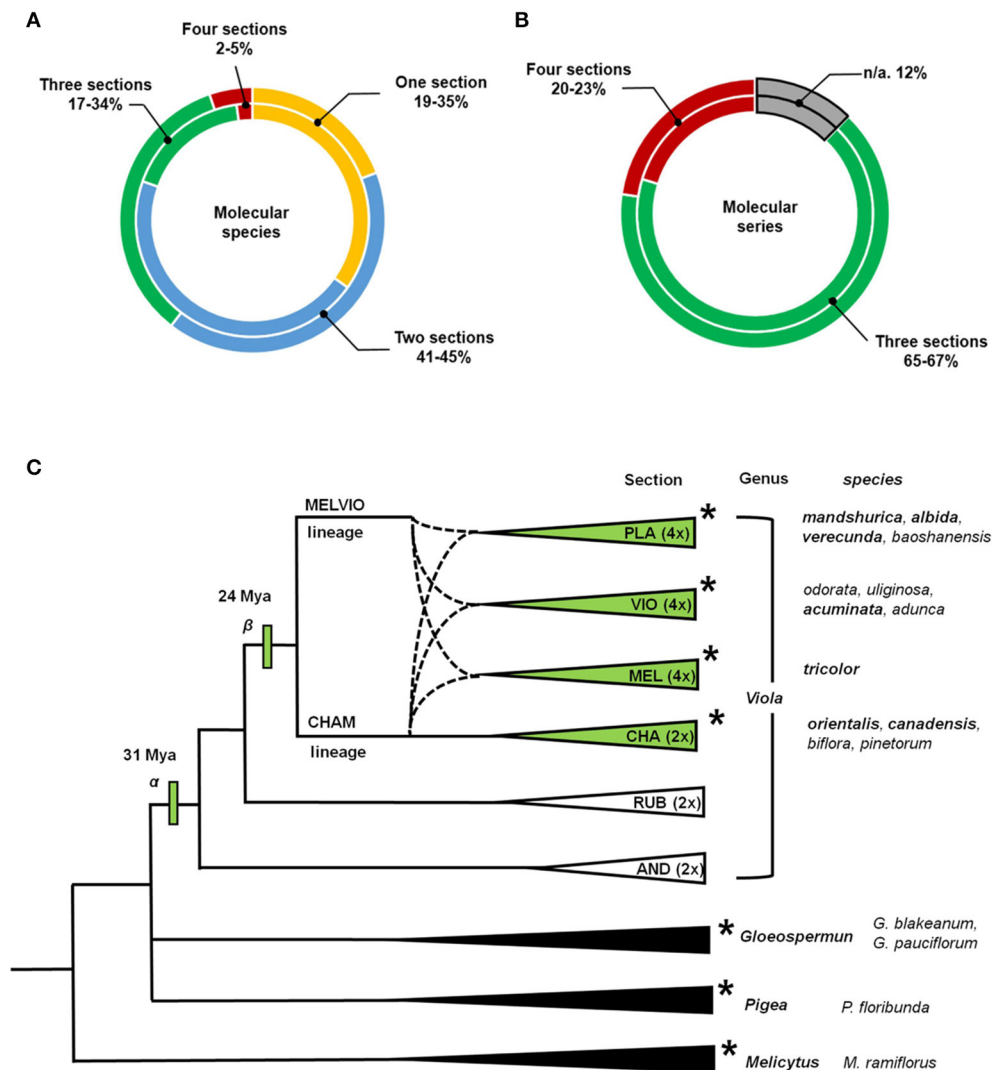


**FIGURE 4 |** Phylogenetic tree of cyclotide precursors. The tree contains 92 precursor sequences in total, with two representative sequences from each of the 46 molecular species. Except from the nine molecular species (i.e., GA2, GP1, NS2, NS5, NL2, HS4, PN1, QD1, and YY1), all other 37 molecular species (80%, 37/46) are grouped as one monophyletic clade (posterior probability ( $pp$ )  $\geq 0.75$ ). Also, except from four molecular series (i.e., FA, HF, HS, and YA), the molecular species belong to other nine molecular series are monophyletic ( $pp \geq 0.97$ ). In the phylogenetic tree, DI1 is a single exception from the lineage classification based on sequence signature. DI1 is grouped into Möbius lineage in the phylogenetic tree ( $pp \geq 0.70$ ), even though it is classified as bracelet lineage based on the sequence signature.

some molecular species that could have occurred by the genetic changes from *Viola* speciation. It should be emphasized that the degree of concurrence is likely higher at the genomic level than

at the transcriptomic level estimated from the current study, because all genomic presence of the molecular species can not be captured by the transcriptome. Some molecular species might not





**FIGURE 5 |** Distribution of molecular species and molecular series over the established phylogenetic relationships of *Viola*. Proportions of the total number of cyclotide molecular species and molecular series found in four different *Viola* sections. **(A)** The values shown in the outer rings shows results obtained by considering numbers of molecular species, and while those in the inner rings show results obtained based on numbers of precursor sequences. **(B)** The values shown in the outer rings shows results obtained by considering numbers of molecular series, and while those in the inner rings show results obtained based on numbers of precursor sequences. Those results are based on data for complete sequences. The result based on both complete and partial sequences is shown in Supplementary Figure 4. **(C)** Phylogenetic relationships between sections and genera of Violaceae used in the analyses conducted within this work. *Viola* Sections are abbreviated as follows: *Plagiostigma* as PLA, *Viola* as VIO, *Melanium* as MEL, *Chamaemelum* as CHA, *Rubellium* as RUB, and *Andinium* as AND. Dotted lines indicate the complex ancestries of the three allotetraploid sections *Melanium*, *Plagiostigma*, and *Viola*, all derived from hybridization between the CHAM and MELVIO lineages 15–20 Ma ago. Genera and sections studied using transcriptomic methods are indicated with asterisks (\*). The total number of species within *Viola* is estimated to be 580–620, most of which (61–65%) belong to these four sections. The first ancestor of the genus ( $\alpha$ ) is dated to 31 Mya, and the common ancestor of the four studied sections ( $\beta$ ) is dated to 24 Mya (Marcussen et al., 2015). The phylogeny of *Viola* is based on the work of Marcussen et al. (2015) and that of Violaceae is based on the work of Wahlert et al. (2014).

be expressed, and the RNA sequence of some molecular species might be degraded during the transcriptomic assay.

Secondly, signature sequences are conserved between *Viola* and other Violaceae genera, i.e., *Gloeospermum*, *Melicytus*, and *Pigea* (Supplementary Figure 5). In particular, the sequence signature of the NTR [−9, −8] region is largely conserved in the Möbius and bracelet lineages: Y<sup>−9</sup>-A<sup>−8</sup> is found in the Möbius lineages of *Viola* and *Melicytus*, G<sup>−9</sup>-A<sup>−8</sup> in the bracelet lineages

of *Viola* and *Gloeospermum*, and H<sup>−9</sup>-S<sup>−8</sup> in the bracelet lineages of *Viola* and *Pigea*. In addition, the sequence signature of the NTPP [−56, −38] is conserved in the bracelet lineage. That region shows high sequence similarity between some precursors, i.e., molecular species NS2 and HS3 from the genus *Viola* and precursors from *Melicytus* and *Pigea*, respectively. Interestingly, the protein sequence of Gpc3, a precursor belonging to the bracelet lineage, is identical in both *Viola* and *Gloeospermum*.



These observations indicate that these sequence signatures have been conserved at least since the divergence of their most recent common ancestor some 50 million years ago (Marcussen, Wahlert et al., in prep.).

Thirdly, the cyclotide sequence diversity of *Viola* is expected to be dependent on the differentiation into sections. It is estimated that 67–88% of the *Viola* sections are allopolyploids (Marcussen et al., 2015) that combine genomes from different diploid lineages. Also, increase in ploidy has been shown experimentally to increase also sequence diversity of cyclotides by mutation, e.g., in *Oldenlandia affinis* (Seydel et al., 2007). However, the cyclotide sequence diversity could not only be very huge by the sequence variations within molecular species, but also be limited by sharing the same molecular species across different sections. The sequence diversity can be large by allowing conservative substitution within the same molecular species. Under the current classification (Marcussen et al., 2015), *Viola* comprises at least 16 sections and some 600 species, of which 10 sections (~454 species, 76%) possess at least one CHAM genome, either alone or, as a result of allopolyploidization, in combination with other CHAM genomes or MELVIO genomes. These 10 sections include the four (~380 species, 63%) inspected by the current study; and the current depth analysis reveals that most of cyclotide precursor sequences across different sections are likely grouped into the molecular species. The precursors identified in the current transcriptome and those reported previously all exhibit striking sequence similarity (Zhang et al., 2009, 2015b; Hellinger et al., 2015; Slazak et al., 2015).

## Expression Profiles of Cyclotides and Their Structural Diversity

The expression profiles of cyclotides were analyzed at transcriptomic and peptidomic levels both. At the transcriptomic level, expression levels of precursor RNA were evaluated by their FPKM values. The expression of those sequences cyclotide sequences were then assayed at the peptidomic level by LC-MS.

At the transcriptomic level, a structurally diverse set of cyclotides were found, including archetypical, linear, hybrid and the novel types of cyclotides described above. In the bracelet lineage, the majority of cyclotide precursors were archetypical bracelets (73%, 125/171). In the Möbius lineage, the proportion of archetypical Möbius (33%, 26/78) was only second largest after the structural hybrids (35%, 28/78). However, judged by FPKM values, the expression levels of archetypical cyclotides were highest within each lineage. Most archetypical precursors have FPKM values higher than 150, i.e., 64% (=9/14) and 79% (=43/54) of the precursors in Möbius lineage and in bracelet lineage, respectively have values larger than the cutoff at 150 (Supplementary Table 4). On the other hand, precursors of novel cyclotides have low FPKM values, i.e., 53% (=7/13) and 100% (=13/13) of the precursors of novel cyclotides in the Möbius lineage and in the bracelet lineage, respectively. Moreover, in the Möbius lineage, 50% (=6/12) of the precursors encoding structural hybrids exceed the FPKM-cutoff value.

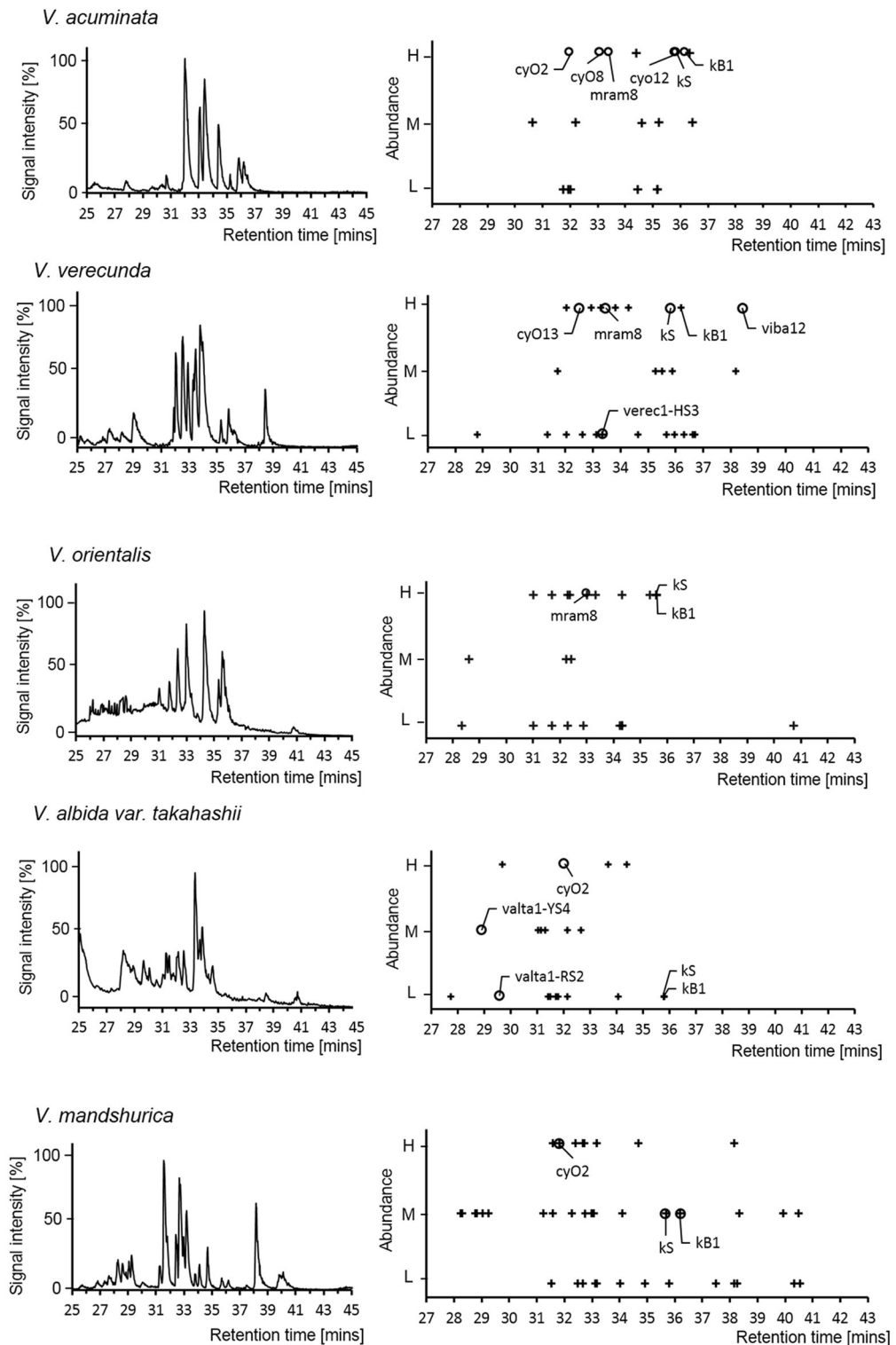
At the peptidomic level, numbers ranging from 19 to 44 cyclotides were detected in each of the *Viola* species. This

number is similar to the number of precursor sequences found at transcriptomic level (23 to 37). Only few of the cyclotides (4–26%) were detected at both levels (Figure 6; see Supplementary Table 5). Most of those cyclotides were archetypical Möbius (kB1 and kS) and archetypical bracelets (cyO2, cyO8, cyO13, mram8, and viba12). Only two of the novel cyclotides were found at both levels, and both were found eluting with very early retention time and at low protein levels of abundance. Cyclotides found at the peptidomic-level only were identified by matching their calculated MW, assuming that they could belong to any type, e.g., be hybrid, archetypical Möbius and bracelet. Discrepancy between expression at transcriptomic- and peptidomic levels has been reported previously (Burman et al., 2010; Hellinger et al., 2015). In those studies cyclotides expressed as peptides were all hybrids, or archetypical Möbius or bracelets. The failure of detection the novel cyclotides using LC-MS could be either from the large hydrophilicity of those novel cyclotides (Supplementary Figure 6) or from their low expression at the peptidomic level. Interestingly, kS (varv A) and kB1 were not found in the transcriptome from two *Viola* species (*V. orientalis* and *V. albida* var. *takahashii*), whereas these particular cyclotides were found at the protein level in all five *Viola* species. Such expression differences could be related to the molecular stability and regulation mechanism. However, it is clear that cyclotides accumulate after their expression in plant tissues.

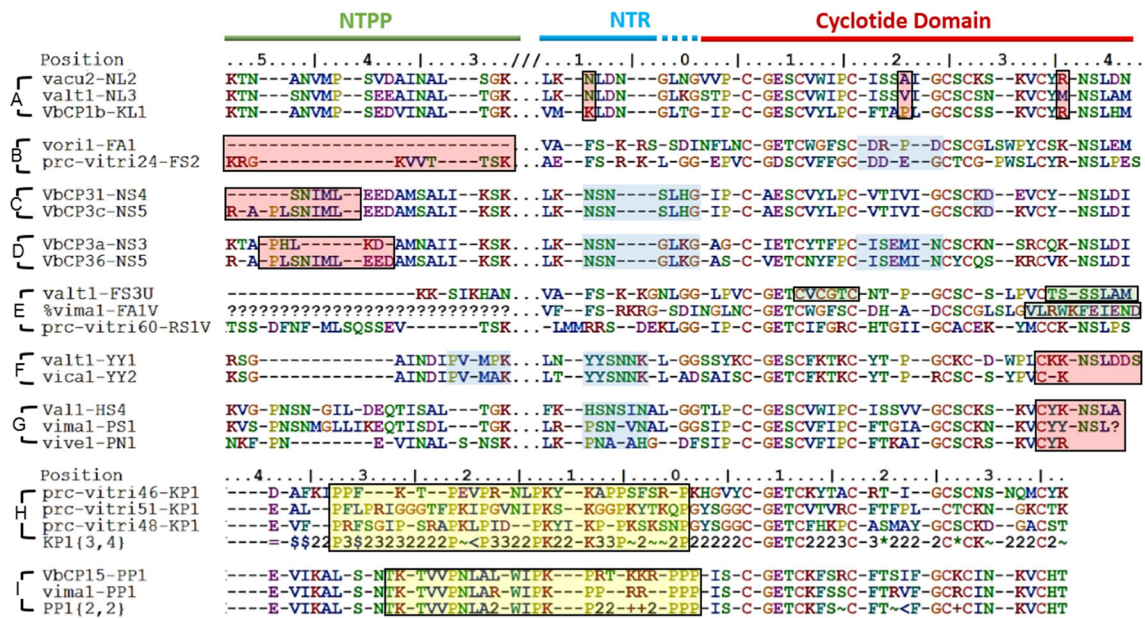
## Major Sequence Changes Are Linked to Neofunctionality

In some cases, the differences between the precursor sequences of closely related molecular species were small (Figure 7A) while in other they were quite large (Figures 7B–D). In this context, a large difference is defined as an indel of more than five residues. With the exception of these indel regions, the precursor sequences are homologous, and the homologous sequence regions contain unique sequence traits found only in those molecular species. This implies that despite their large sequence differences, those molecular species are closely related. Such an outcome is analogous to that predicted by the theory of punctuated equilibrium (Eldredge and Gould, 1997), which describes evolutionary trends and speciation at the organism level. Another possible explanation for the observations is that mutations gradually accumulate in cyclotide precursors, leading to the loss or inactivation of intermediate sequences.

Some of the cyclotide domains exhibited unusual sequences, which fell into three main groups (Figure 7E). The first such group contained sequences whose cyclotide domains have an uneven number of cysteines (five or seven). Having uneven number of cysteines makes the formation of a cystine knot impossible, and can create the possibility of forming disulfide bridges not associated with the cystine knot, as well as the potential for dimerization. The second group consists of sequences that lack a residue required for AEP-mediated cyclisation, i.e., an N or D residue in loop 6. Abnormal precursor sequences of this sort are presumably the result of mutations in the genes encoding cyclic cyclotide precursors rather than being inherited from an ancestral linear cyclotide because they exhibit



**FIGURE 6 |** Cyclotide expression on peptidomic level. LC-MS chromatogram of the five *Viola* species used for transcriptome assays in current study (left panel). The abundance of cyclotides is linked to their signal intensity (Right panel). For each *Viola* species, the LC-MS chromatogram (Base Peak Ion) and mass distribution of the cyclotides are shown with retention times and abundance levels (High/Medium/Low). The symbols + and o denote cyclotides that were not found in the transcriptome and were found in the transcriptome, respectively. The cyclotides (kB1, kS, cyO2, cyO8, and cyO12) were confirmed by comparing the retention time and isotopic mass derived from the LC-MS of *V. tricolor* and *V. odorata* as reference chromatogram.



**FIGURE 7 |** Sequence diversity between closely related molecular species. Some molecular species differ by only a few residues, e.g., (A) while others exhibit pronounced sequence variation, e.g., (B–D) Sequence differences are highlighted by red boxes. Common sequence traits (indicated by blue boxes) are found in closely related precursors. (E) Differences in the residues forming the CCK structural motif (green boxes). Additional cysteines are found in loop 2 of valt1-FS3U and loop 6 of prc-vitri60-RS1V. The cyclisation-enabling N/D residues are absent in loop 6 of valt1-FS3U. Vima1-FA1V lacks a cysteine residue in loop 6, which is elongated. (F) and (G). Emergence of a cyclic backbone in the Möbius and bracelet lineages, respectively. Both lineages exhibit large sequence variation in the C-terminal loop 6. (H,I) Proline-rich sequence regions (indicated by yellow boxes).

strong sequence similarity with the former class of precursors in both loop 6 and the C-terminal tail. Conversely, typical linear cyclotide precursors have short loop 6 sequences, lack a C-terminal tail, and are quite abundant. The third group of abnormal precursor sequences included sequences with one or more unusually short or long loops. Only five such sequences were identified (among a total of 283), but the three types of abnormality often co-occurred in the same precursor. For example, valt1-FS3U contains one additional cysteine in loop 2, and lacks an N/D residue in loop 6. All of these precursors are only weakly expressed.

The cyclic backbone seems to have emerged by molecular speciation. Comparing sequences from linear and cyclic molecular species, i.e., the cyclic YY1 and linear YY2 in Möbius lineage, and the cyclic PS1 and HS4 to the linear PN1 in the bracelet lineage (Figures 7F,G), suggests that linearity is likely to be the primitive (ancestral) trait, rather than linear cyclotides having evolved from cyclic ancestors via the mutational loss of the N/D residue in loop 6 that is needed for cyclisation. Although the cyclotide precursor sequences are too variable at the nucleotide level to resolve the relationships between linear and bracelet (Huson and Bryant, 2006), linearity being ancestral is supported by three facts. First, linear cyclotides that have undergone such mutational losses normally have elongated loop 6 sequences with mutations in the C-terminal tail. Second, the prodomain sequences of these molecular species are well-defined (all of them share the same sequence signature). Third, the

molecular species corresponding to these putative primitive linear cyclotides are found in all four studied sections of *Viola* and also in other genera (Mra13).

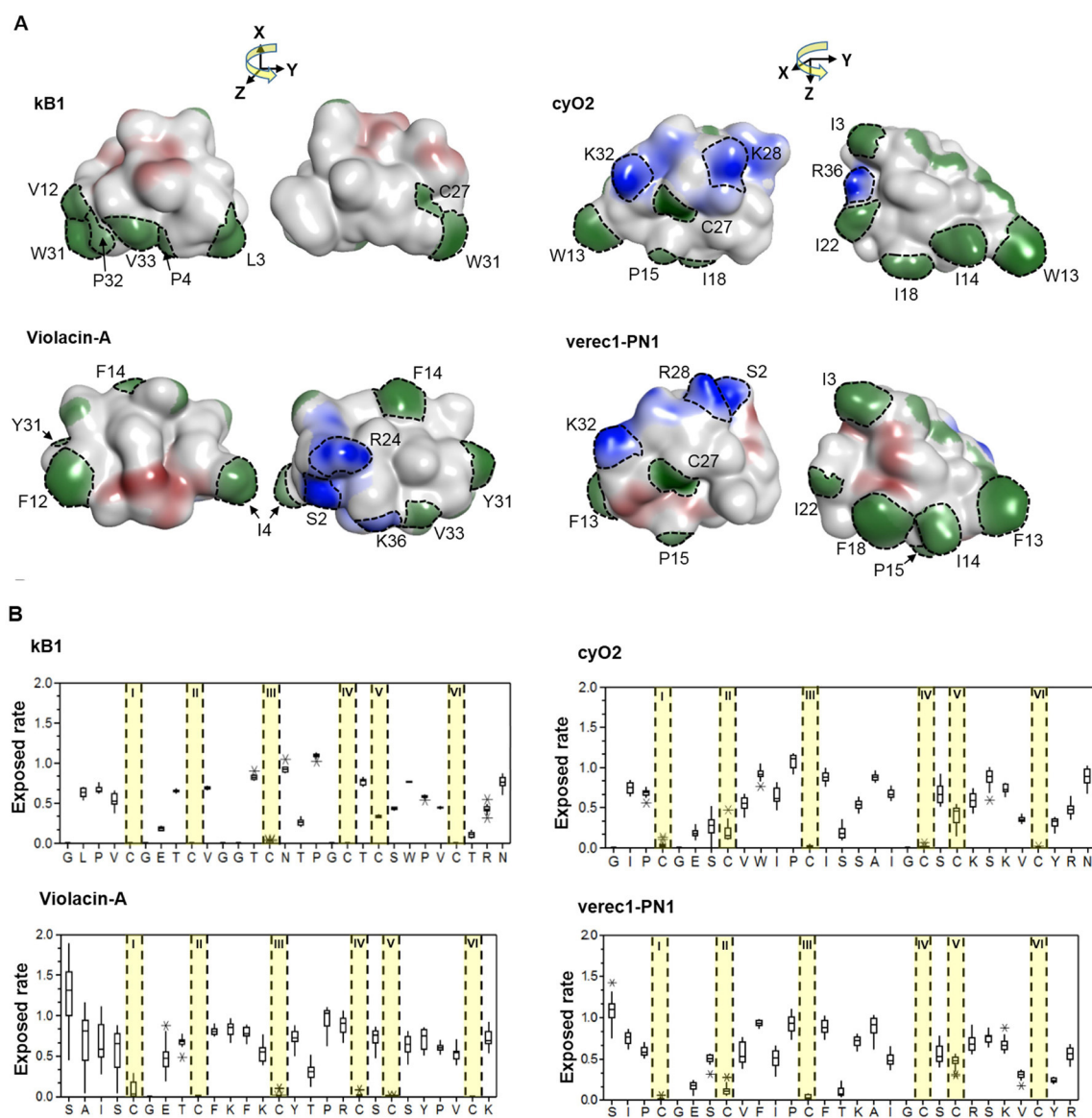
Some precursors sequences could also be grouped as rich in prolines (Figures 7H,I). The common structure-based criterion used to identify mature Möbius cyclotides is the presence of a *cis* Pro in loop 5. Classification based on this approach is not generally consistent with the evolutionary classification into lineages: many precursor sequences assigned to the Möbius lineage lack this Pro residue in loop 5 (P32) of the cyclotide domain. Examples include the molecular species YA1, HF1, and YS1-3, which exhibit appreciable sequence similarity with archetypical Möbius cyclotides in loops 2 and 3. These sequence intermediates, which have characteristics of both bracelet and Möbius cyclotides, have been classified as the hybrid subfamily, and it has been suggested that they are genetic chimeras of bracelets and Möbius cyclotides (Daly et al., 2006). Our analysis suggests that this is not the case, at least in the Violaceae. Instead, these structural hybrids and archetypical Möbius cyclotides seem to originate from the molecular species YY1 of the Möbius lineage, which in turn is closely related to the linear ancestor YY2.

A key question is whether the emergence of the cyclic backbone changed the properties of the membrane-interacting surfaces of the linear cyclotide sequences. To answer this question, we analyzed the structural traits (i.e., the exposed ratio and physicochemical properties) of the residues that



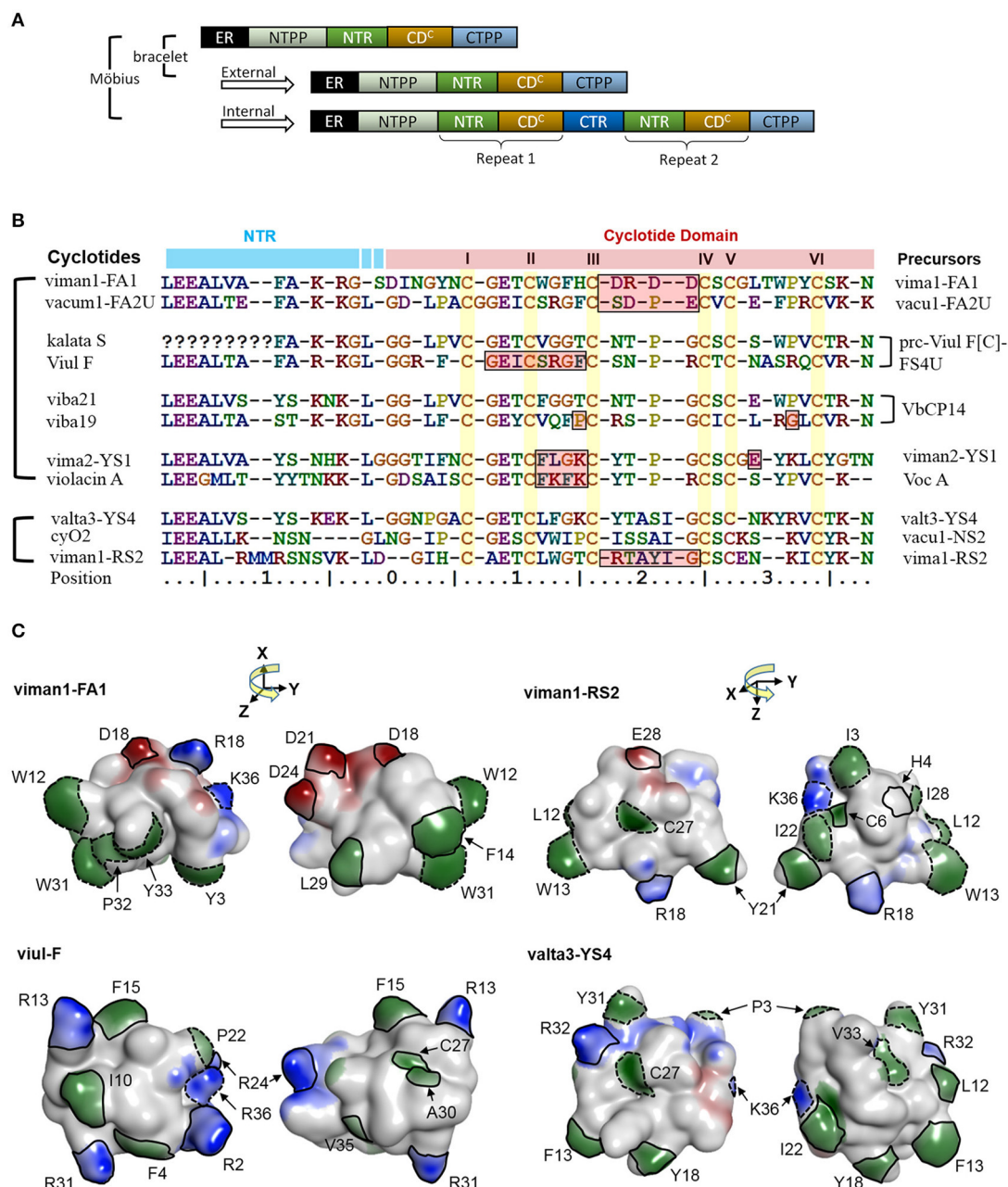
differ between linear and cyclic cyclotides within Möbius and bracelet lineages. The lineages were considered separately, so the cyclic Möbius cyclotide kalata B1 was compared to the linear cyclotide violacin A from the Möbius lineage. Similarly, the cyclic bracelet cyclotide cyO2 was compared to the linear verec1-PN1 from the bracelet lineage. In both cases, the linear cyclotide appeared to be ancestral to its cyclic counterpart. Linear and cyclic cyclotides from the same lineage exhibit substantial

sequence similarity, with extensive conservation of residues' physicochemical properties at key positions. This is illustrated by the example of the residues in loop 6 (**Figure 8B**). In the linear cyclotides, loop 6 is conformationally flexible, and the charged groups of the termini are largely exposed on the molecular surface (**Figures 8A,B**). The exposed ratios and physicochemical properties of the corresponding residues in the cyclic cyclotides are very similar to those in their linear counterparts, implying



**FIGURE 8 |** Membrane-interactive properties of linear and cyclic cyclotides within the Möbius and bracelet lineages. **(A)** Structures of kalata B1 and violacin A from the Möbius lineage, and cyO2 and verec1-PN1 from the bracelet lineage. Surface-exposed side-chains are shown in blue if positively charged, red if negatively charged, and green if hydrophobic. Notably, the N-terminal residues (S2) of the linear cyclotides always give rise to an exposed positive charge on the molecular surface. The numbering of residues follow the consensus sequence of the cyclotide domain in **Figure 3**. **(B)** Exposed ratios of linear and cyclic cyclotides' residues within each lineage. The exposure ratio compares the solvent-exposed surface area (SASA) of each residue in a protein to its SASA in the tripeptide Gly-X-Gly ( $\psi = \phi = 180^\circ$ ); the more exposed the residue is, the greater the value. With the exception of the residues in loop 6, residues of cyclic and linear cyclotides have similar exposure ratios at corresponding positions. For each residue, the boxes show the range between the first and third quartiles, and the upper and lower error bars represent the minimum and maximum exposure ratios, respectively.





**FIGURE 9 |** A potential link between precursor architecture and neofunctionality. **(A)** Suggested modes of duplication in precursor genes. In *Viola*, the precursor architecture differs between lineages. Precursors of bracelet lineages appear to contain only one modular domain (i.e., a concatenated NTR and a cyclotide domain), whereas precursors of the Möbius lineage may contain one to three such domains. The repeats in the Möbius lineage probably originated from internal gene duplication events. All precursors sharing the same architecture, in both the Möbius and bracelet lineages, probably originated from external gene duplication. **(B)** Sequence alignment between new types of cyclotides and previously identified archetypes. In the Möbius lineage, sequences vary between repeated cyclotide domains within the same precursor and cyclotide domains of different precursors. There are high levels of sequence conservation within repeated cyclotide domains, as exemplified by kB1 and kS within Voc1. However, the sequences of cyclotide domains in precursors containing multiple cyclotide domains in general are frequently very different, as exemplified by kalata S and viul F within prec-Viul F[C]-FS4U, and viba 19 and viba 21 within VbCP14. Of the precursor sequences shown in the figure, the molecular species YS7, which belongs to the Möbius lineage, has a similar sequence to archetypal bracelet cyclotides. For example, valta3-YS4 lacks a proline residue in loop 5 and has an elongated loop 3. Conversely, the molecular species FA1 contains lipophilic residues in loops 2 and 5, while loop 3 consists mainly of negatively charged residues. **(C)** Comparison of cyclotides' surfaces and electrostatic potentials. New types of cyclotides — viman-FA1, viman1-RS2, and viul-F — are shown with their electrostatic potential surfaces. The distributions of their electrostatic potentials are generally dissimilar to those seen in typical bracelet and Möbius cyclotides. Residues with high exposed ratios are indicated by solid and dashed lines on the molecular surfaces. Residues highlighted with solid lines have dissimilar physicochemical properties to those found in the corresponding positions of archetypal cyclotides. Residues whose physicochemical properties match those of the archetypal cyclotides are indicated by dotted lines. Residues are numbered in accordance with their sequence positions in **(B)**. Negatively charged surface regions are colored in red, positively charged regions in blue, and hydrophobic regions in green. The numbering of residues follow the consensus sequence of the cyclotide domain in **Figure 3**.

that the structural framework favoring membrane interaction emerged before the cyclic backbone.

Some of the cyclotide domains exhibit very high levels of sequence diversity between different molecular species. Some molecular species (e.g., YS1-4, RS2 and FA1) contain cyclotide domain sequences that are highly dissimilar to archetypal cyclotide sequences (Figure 9). We found that these molecular species occur across *Viola* sections, which means that they must have evolved before the differentiation of these sections some 15 million years ago. Although the biological function of all cyclotide forms is currently unclear, their maintenance within the genome suggests that they fulfill vital and divergent functions in their host plants.

It is likely that some cyclotides may have biological functions based on mechanisms other than membrane binding and disruption (Burman et al., 2011). For example, the cyclotides-FA1 might have different orientations to archetypal cyclotides, even when one compares them to others from the same lineage. They have lipophilic residues in loops 2 and 5 with large exposed areas on the molecular surface. The lipophilicity of this loop differs from that of archetypal cyclotides: the archetypal Möbius cyclotides do not have lipophilic residues in loop 2, and the archetypal bracelets do not have lipophilic residues in loop 5. In addition, loop 3 of the FA1 cyclotides is rich in negatively charged residues, making them unique among the molecular species found in the genus *Viola* and suggesting that their functions may be unrelated to membrane binding.

Contradicting previous hypotheses, it appears that the cyclotides of the Möbius lineage exhibit a somewhat greater diversity of sequence traits than those of the bracelet lineage. Such large sequence diversity of the Möbius lineage may be correlated with the combined occurrence of internal and external gene duplications in the cyclotide domain. Precursor architecture with such duplication of the cyclotide domain is illustrated in the Figure 9A. A large proportion of the precursors (72%; 56/78) belonging to the Möbius lineage contain multiple cyclotide domains that appear to have originated from internal duplication events. Conversely, no bracelet precursors having repeated

cyclotide domains have yet been identified in the Violaceae. There are large sequence differences in certain precursors' cyclotide domains at positions close to the N-terminal prodomains, as can be seen by comparing the sequences of the FA1 and YS4 molecular species to the archetypal Möbius cyclotides varv A and kalata S (Figures 9B,C). Pronounced sequence differences also exist between repeated cyclotide domains in some individual precursors, as in the case of the precursor of Viul F. Together with the rest of our findings, this implies that combined internal and external duplications can synergistically produce large changes in cyclotides' sequences and structures, giving rise to new biological functions: neofunctionality.

## AUTHOR CONTRIBUTIONS

SP, TM, KJR, ID, AB, and UG designed experiments. SP carried out interpretation of precursor sequences, modeling and mass spectrometry. UG supervised the experiments. K-OY collected plant material for the transcriptome sequencing. EJ identified cyclotide genes from transcriptomic data. TM and SP performed the phylogenetic analyses. All authors contributed to the writing and revision of the manuscript.

## FUNDING

This research was supported by Swedish Research Council (#2012-5063). KJR is an ARC Future Fellow (FT130100890).

## ACKNOWLEDGMENTS

We are thankful to Dr. Kyeong-Sik Cheon at Kangwon National University for valuable discussions.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2017.02058/full#supplementary-material>

## REFERENCES

- Arnison, P. G., Bibb, M. J., Bierbaum, G., Bowers, A. A., Bugni, T. S., Bulaj, G., et al. (2013). Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* 30, 108–160. doi: 10.1039/C2NP20085F
- Ballard, H. E., Sytsma, K. J., and Kowal, R. R. (1998). Shrinking the violets: phylogenetic relationships of infrageneric groups in *Viola* (Violaceae) based on internal transcribed spacer DNA sequences. *Syst. Bot.* 23, 439–458. doi: 10.2307/2419376
- 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
- Burman, R., Gruber, C. W., Rizzardi, K., Herrmann, A., Craik, D. J., Gupta, M. P., et al. (2010). Cyclotide proteins and precursors from the genus *Gloeospermum*: filling a blank spot in the cyclotide map of Violaceae. *Phytochemistry* 71, 13–20. doi: 10.1016/j.phytochem.2009.09.023
- Burman, R., Gunasekera, S., Strömstedt, A. A., and Göransson, U. (2014). Chemistry and biology of cyclotides: circular plant peptides outside the box. *J. Nat. Prod.* 77, 724–736. doi: 10.1021/np401055j
- Burman, R., Herrmann, A., Tran, R., Kivelä, J. E., Lomize, A., Gullbo, J., et al. (2011). Cytotoxic potency of small macrocyclic knot proteins: structure-activity and mechanistic studies of native and chemically modified cyclotides. *Org. Biomol. Chem.* 9, 4306–4314. doi: 10.1039/c0ob00966k
- Burman, R., Yeshak, M. Y., Larsson, S., Craik, D. J., Rosengren, K. J., and Göransson, U. (2015). Distribution of circular proteins in plants: large-scale mapping of cyclotides in the Violaceae. *Front. Plant Sci.* 6:855. doi: 10.3389/fpls.2015.00855
- Claeson, P., Göransson, U., Johansson, S., Luijendijk, T., and Bohlin, L. (1998). Fractionation protocol for the isolation of polypeptides from plant biomass. *J. Nat. Prod.* 61, 77–81. doi: 10.1021/np970342r
- Colgrave, M. L., Kotze, A. C., Huang, Y. H., O'Grady, J., Simonsen, S. M., and Craik, D. J. (2008a). Cyclotides: natural, circular plant peptides that possess significant activity against gastrointestinal nematode parasites of sheep. *Biochemistry* 47, 5581–5589. doi: 10.1021/bi800223y
- Colgrave, M. L., Kotze, A. C., Ireland, D. C., Wang, C. K., and Craik, D. J. (2008b). The anthelmintic activity of the cyclotides: natural variants

- with enhanced activity. *ChemBiochem* 9, 1939–1945. doi: 10.1002/cbic.200800174
- Craik, D. J., Daly, N. L., Bond, T., and Waite, C. (1999). Plant cyclotides: a unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif. *J. Mol. Biol.* 294, 1327–1336. doi: 10.1006/jmbi.1999.3383
- Daly, N. L., Clark, R. J., Plan, M. R., and Craik, D. J. (2006). Kalata B8, a novel antiviral circular protein, exhibits conformational flexibility in the cystine knot motif. *Biochem. J.* 393(Pt 3), 619–626. doi: 10.1042/BJ20051371
- 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
- Drummond, A. J., and Rambaut, A. (2007). BEAST: bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214. doi: 10.1186/1471-2148-7-214
- Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29, 1969–1973. doi: 10.1093/molbev/mss075
- Dutton, J. L., Renda, R. F., Waite, C., Clark, R. J., Daly, N. L., Jennings, C. V., et al. (2004). Conserved structural and sequence elements implicated in the processing of gene-encoded circular proteins. *J. Biol. Chem.* 279, 46858–46867. doi: 10.1074/jbc.M407421200
- Eldredge, N., and Gould, S. J. (1997). On punctuated equilibria. *Science* 276, 338–341. doi: 10.1126/science.276.5311.337c
- Görransson, U., Sjögren, M., Svängård, E., Claesson, P., and Bohlin, L. (2004). Reversible antifouling effect of the cyclotide cycloviolacin O<sub>2</sub> against barnacles. *J. Nat. Prod.* 67, 1287–1290. doi: 10.1021/np0499719
- Gran, L. (1973a). On the effect of a polypeptide isolated from “Kalata-Kalata” (*Oldenlandia affinis* DC) on the oestrogen dominated uterus. *Acta Pharmacol. Toxicol. (Copenh.)* 33, 400–408. doi: 10.1111/j.1600-0773.1973.tb01541.x
- Gran, L. (1973b). Oxytocic principles of *Oldenlandia affinis*. *Lloydia* 36, 174–178.
- Gruber, C. W., Elliott, A. G., Ireland, D. C., Delprete, P. G., Dessein, S., Görransson, U., et al. (2008). Distribution and evolution of circular miniproteins in flowering plants. *Plant Cell* 20, 2471–2483. doi: 10.1105/tpc.108.062331
- Gründemann, C., Koehbach, J., Huber, R., and Gruber, C. W. (2012). Do plant cyclotides have potential as immunosuppressant peptides? *J. Nat. Prod.* 75, 167–174. doi: 10.1021/np200722w
- Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704. doi: 10.1080/10635150390235520
- Gustafson, K. R., Sowder, R. C., Henderson, L. E., Parsons, I. C., Kashman, Y., Cardellina, J. H., et al. (1994). Circulin-A and B: novel HIV-inhibitory macrocyclic peptides from the tropical tree *Chassalia parvifolia*. *J. Am. Chem. Soc.* 116, 9337–9338. doi: 10.1021/ja00099a064
- Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Series* 41, 95–98.
- Harris, K. S., Durek, T., Kaas, Q., Poth, A. G., Gilding, E. K., Conlan, B. F., et al. (2015). Efficient backbone cyclization of linear peptides by a recombinant asparaginyl endopeptidase. *Nat. Commun.* 6:10199. doi: 10.1038/ncomms10199
- Hellinger, R., Koehbach, J., Soltis, D. E., Carpenter, E. J., Wong, G. K., and Gruber, C. W. (2015). Peptidomics of circular cysteine-rich plant peptides: analysis of the diversity of cyclotides from *Viola tricolor* by transcriptome and proteome mining. *J. Proteome Res.* 14, 4851–4862. doi: 10.1021/acs.jproteome.5b00681
- Henriques, S. T., Huang, Y. H., Rosengren, K. J., Franquelin, H. G., Carvalho, F. A., Johnson, A., et al. (2011). Decoding the membrane activity of the cyclotide kalata B1: the importance of phosphatidylethanolamine phospholipids and lipid organization on hemolytic and anti-HIV activities. *J. Biol. Chem.* 286, 24231–24241. doi: 10.1074/jbc.M111.253393
- Hernandez, J. F., Gagnon, J., Chiche, L., Nguyen, T. M., Andrieu, J. P., Heitz, A., et al. (2000). Squash trypsin inhibitors from *Momordica cochinchinensis* exhibit an atypical macrocyclic structure. *Biochemistry* 39, 5722–5730. doi: 10.1021/bi9929756
- Herrmann, A., Burman, R., Mylne, J. S., Karlsson, G., Gullbo, J., Craik, D. J., et al. (2008). The alpine violet, *Viola biflora*, is a rich source of cyclotides with potent cytotoxicity. *Phytochemistry* 69, 939–952. doi: 10.1016/j.phytochem.2007.10.023
- Huson, D. H., and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23, 254–267. doi: 10.1093/molbev/msj030
- Ireland, D. C., Colgrave, M. L., Nguyencong, P., Daly, N. L., and Craik, D. J. (2006). Discovery and characterization of a linear cyclotide from *Viola odorata*: implications for the processing of circular proteins. *J. Mol. Biol.* 357, 1522–1535. doi: 10.1016/j.jmb.2006.01.051
- Jennings, C., West, J., Waite, C., Craik, D., and Anderson, M. (2001). Biosynthesis and insecticidal properties of plant cyclotides: the cyclic knotted proteins from *Oldenlandia affinis*. *Proc. Natl. Acad. Sci. U.S.A.* 98, 10614–10619. doi: 10.1073/pnas.191366898
- Kaas, Q., and Craik, D. J. (2010). Analysis and classification of circular proteins in CyBase. *Biopolymers* 94, 584–591. doi: 10.1002/bip.21424
- Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323. doi: 10.1186/1471-2105-12-323
- Lindholm, P., Görransson, U., Johansson, S., Claesson, P., Gullbo, J., Larsson, R., et al. (2002). Cyclotides: a novel type of cytotoxic agents. *Mol. Cancer Ther.* 1, 365–369. Available online at: <http://mct.aacrjournals.org/content/1/6/365>
- Marcussen, T., Heier, L., Brysting, A. K., Oxelman, B., and Jakobsen, K. S. (2015). From gene trees to a dated allopolyploid network: insights from the angiosperm genus *Viola* (Violaceae). *Syst. Biol.* 64, 84–101. doi: 10.1093/sysbio/syu071
- Marcussen, T., Jakobsen, K. S., Danihelka, J., Ballard, H. E., Blaxland, K., Brysting, A. K., et al. (2012). Inferring species networks from gene trees in high-polyploid North American and Hawaiian violets (*Viola*, Violaceae). *Syst. Biol.* 61, 107–126. doi: 10.1093/sysbio/syr096
- Mulvenna, J. P., Mylne, J. S., Bharathi, R., Burton, R. A., Shirley, N. J., Fincher, G. B., et al. (2006). Discovery of cyclotide-like protein sequences in graminaceous crop plants: ancestral precursors of circular proteins? *Plant Cell* 18, 2134–2144. doi: 10.1105/tpc.106.042812
- Mulvenna, J. P., Sando, L., and Craik, D. J. (2005). Processing of a 22 kDa precursor protein to produce the circular protein tricyclon A. *Structure* 13, 691–701. doi: 10.1016/j.str.2005.02.013
- Mylne, J. S., Chan, L. Y., Chanson, A. H., Daly, N. L., Schaefer, H., Bailey, T. L., et al. (2012). Cyclic peptides arising by evolutionary parallelism via asparaginyl-endopeptidase-mediated biosynthesis. *Plant Cell* 24, 2765–2778. doi: 10.1105/tpc.112.099085
- Nguyen, G. K., Lian, Y., Pang, E. W., Nguyen, P. Q., Tran, T. D., and Tam, J. P. (2013). Discovery of linear cyclotides in monocot plant *Panicum laxum* of Poaceae family provides new insights into evolution and distribution of cyclotides in plants. *J. Biol. Chem.* 288, 3370–3380. doi: 10.1074/jbc.M112.415356
- Nguyen, G. K., Zhang, S., Nguyen, N. T., Nguyen, P. Q., Chiu, M. S., Hardjojo, A., et al. (2011). Discovery and characterization of novel cyclotides originated from chimeric precursors consisting of albumin-1 chain a and cyclotide domains in the Fabaceae family. *J. Biol. Chem.* 286, 24275–24287. doi: 10.1074/jbc.M111.229922
- Northfield, S. E., Wang, C. K., Schroeder, C. I., Durek, T., Kan, M. W., Swedberg, J. E., et al. (2014). Disulfide-rich macrocyclic peptides as templates in drug design. *Eur. J. Med. Chem.* 77, 248–257. doi: 10.1016/j.ejmech.2014.03.011
- Okonechnikov, K., Golosova, O., Fursov, M., and team, U. (2012). Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* 28, 1166–1167. doi: 10.1093/bioinformatics/bts091
- Park, S., Strömstedt, A. A., and Görransson, U. (2014). Cyclotide structure-activity relationships: qualitative and quantitative approaches linking cytotoxic and anthelmintic activity to the clustering of physicochemical forces. *PLoS ONE* 9:e91430. doi: 10.1371/journal.pone.0091430
- Plan, M. R., Saska, I., Cagauan, A. G., and Craik, D. J. (2008). Backbone cyclized peptides from plants show molluscicidal activity against the rice pest *Pomacea canaliculata* (golden apple snail). *J. Agric. Food Chem.* 56, 5237–5241. doi: 10.1021/jf800302f
- Poth, A. G., Colgrave, M. L., Lyons, R. E., Daly, N. L., and Craik, D. J. (2011a). Discovery of an unusual biosynthetic origin for circular proteins in legumes. *Proc. Natl. Acad. Sci. U.S.A.* 108, 10127–10132. doi: 10.1073/pnas.1103660108
- Poth, A. G., Colgrave, M. L., Philip, R., Kerenga, B., Daly, N. L., Anderson, M. A., et al. (2011b). Discovery of cyclotides in the fabaceae plant family provides new insights into the cyclization, evolution, and distribution of circular proteins. *ACS Chem. Biol.* 6, 345–355. doi: 10.1021/cb100388j
- Poth, A. G., Mylne, J. S., Grassl, J., Lyons, R. E., Millar, A. H., Colgrave, M. L., et al. (2012). Cyclotides associate with leaf vasculature and are the products

- of a novel precursor in *Petunia* (Solanaceae). *J. Biol. Chem.* 287, 27033–27046. doi: 10.1074/jbc.M112.370841
- Pränting, M., Lööv, C., Burman, R., Göransson, U., and Andersson, D. I. (2010). The cyclotide cycloviolacin O<sub>2</sub> from *Viola odorata* has potent bactericidal activity against Gram-negative bacteria. *J. Antimicrob. Chemother.* 65, 1964–1971. doi: 10.1093/jac/dkq220
- Ravipati, A. S., Henriques, S. T., Poth, A. G., Kaas, Q., Wang, C. K., Colgrave, M. L., et al. (2015). Lysine-rich cyclotides: a new subclass of circular knotted proteins from violaceae. *ACS Chem. Biol.* 10, 2491–2500. doi: 10.1021/acschembio.5b00454
- Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. *Trends Genet.* 16, 276–277. doi: 10.1016/S0168-9525(00)00204-2
- Schöpfke, T., Hasan, A. M. I., Kraft, R., Otto, A., and Hiller, K. (1993). Haemolytisch aktive Komponenten aus *Viola tricolor* L. und *Viola arvensis* murray. *Sci. Pharm.* 61, 145–153.
- Seydel, P., Gruber, C. W., Craik, D. J., and Dörnenburg, H. (2007). Formation of cyclotides and variations in cyclotide expression in *Oldenlandia affinis* suspension cultures. *Appl. Microbiol. Biotechnol.* 77, 275–284. doi: 10.1007/s00253-007-1159-6
- Sigrist, C. J., Cerutti, L., Hulo, N., Gattiker, A., Falquet, L., Pagni, M., et al. (2002). PROSITE: a documented database using patterns and profiles as motif descriptors. *Brief. Bioinformatics* 3, 265–274. doi: 10.1093/bib/3.3.265
- Simonsen, S. M., Sando, L., Ireland, D. C., Colgrave, M. L., Bharathi, R., Göransson, U., et al. (2005). A continent of plant defense peptide diversity: cyclotides in Australian *Hybanthus* (Violaceae). *Plant Cell* 17, 3176–3189. doi: 10.1105/tpc.105.034678
- Simonsen, S. M., Sando, L., Rosengren, K. J., Wang, C. K., Colgrave, M. L., Daly, N. L., et al. (2008). Alanine scanning mutagenesis of the prototypic cyclotide reveals a cluster of residues essential for bioactivity. *J. Biol. Chem.* 283, 9805–9813. doi: 10.1074/jbc.M709303200
- Slazak, B., Jacobsson, E., Kuta, E., and Göransson, U. (2015). Exogenous plant hormones and cyclotide expression in *Viola uliginosa* (Violaceae). *Phytochemistry* 117, 527–536. doi: 10.1016/j.phytochem.2015.07.016
- Tam, J. P., Lu, Y. A., Yang, J. L., and Chiu, K. W. (1999). An unusual structural motif of antimicrobial peptides containing end-to-end macrocycle and cystine-knot disulfides. *Proc. Natl. Acad. Sci. U.S.A.* 96, 8913–8918. doi: 10.1073/pnas.96.16.8913
- Tokuoka, T. (2008). Molecular phylogenetic analysis of Violaceae (Malpighiales) based on plastid and nuclear DNA sequences. *J. Plant Res.* 121, 253–260. doi: 10.1007/s10265-008-0153-0
- Trabi, M., and Craik, D. J. (2002). Circular proteins-no end in sight. *Trends Biochem. Sci.* 27, 132–138. doi: 10.1016/S0968-0004(02)02057-1
- Trabi, M., Mylne, J. S., Sando, L., and Craik, D. J. (2009). Circular proteins from *Melicytus* (Violaceae) refine the conserved protein and gene architecture of cyclotides. *Org. Biomol. Chem.* 7, 2378–2388. doi: 10.1039/b823020j
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt.1621
- Wahlert, G. A., Marcussen, T., Paula-Souza, J., and Ballard, H. E. Jr. (2014). A phylogeny of the Violaceae (Malpighiales) inferred from plastid DNA sequences: implications for generic diversity and intrafamilial classification. *Syst. Bot.* 39, 239–252. doi: 10.1600/036364414X678008
- Witherup, K. M., Bogusky, M. J., Anderson, P. S., Ramjit, H., Ransom, R. W., Wood, T., et al. (1994). Cyclopsychotride, A., a biologically active, 31-residue cyclic peptide isolated from *Psychotria longipes*. *J. Nat. Prod.* 57, 1619–1625. doi: 10.1021/np50114a002
- Yockteng, R. H. E. B. J., G., Mansion, I., and Dajoz, S., Nadot (2003). Relationships among pansies (*Viola* section *Melanium*) investigated using ITS and ISSR markers. *Plant Syst. Evol.* 241, 153–170. doi: 10.1007/s00606-003-0045-7
- Zhang, J., Hua, Z., Huang, Z., Chen, Q., Long, Q., Craik, D. J., et al. (2015a). Two Blast-independent tools, CyPerl and CyExcel, for harvesting hundreds of novel cyclotides and analogues from plant genomes and protein databases. *Planta* 241, 929–940. doi: 10.1007/s00425-014-2229-5
- Zhang, J., Liao, B., Craik, D. J., Li, J. T., Hu, M., and Shu, W. S. (2009). Identification of two suites of cyclotide precursor genes from metallophyte *Viola baoshanensis*: cDNA sequence variation, alternative RNA splicing and potential cyclotide diversity. *Gene* 431, 23–32. doi: 10.1016/j.gene.2008.11.005
- Zhang, J., Li, J., Huang, Z., Yang, B., Zhang, X., Li, D., et al. (2015b). Transcriptomic screening for cyclotides and other cysteine-rich proteins in the metallophyte *Viola baoshanensis*. *J. Plant Physiol.* 178, 17–26. doi: 10.1016/j.jplph.2015.01.017
- Zhu, S., Darbon, H., Dyason, K., Verdonck, F., and Tytgat, J. (2003). Evolutionary origin of inhibitor cystine knot peptides. *FASEB J.* 17, 1765–1767. doi: 10.1096/fj.02-1044fje

**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 © 2017 Park, Yoo, Marcussen, Backlund, Jacobsson, Rosengren, Doo and Göransson. 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) or licensor 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 Challenge to Translate OMICS Data to Whole Plant Physiology: The Context Matters

Marcelo N. do Amaral and Gustavo M. Souza\*

Department of Botany, Institute of Biology, Federal University of Pelotas, Pelotas, Brazil

**Keywords:** downward causation, emergent properties, hierarchical systems, plant signaling, systems biology

## INTRODUCTION: SOME CHALLENGES

The exponential development of high-throughput technologies in the last decades, supporting and improving the OMICS science, has allowed uncovering successfully the complexity of the organizational network patterns in the cell's metabolism to the plant phenome, founding the science of system biology (Mochida and Shinozaki, 2011). Further, the huge data sets and growing computational power have stimulated scientists to glimpse about how plants respond to the environmental changes, and how such knowledge could engender new technologies, for instance, to increase crop yields (Edwards and Batley, 2004; Tardieu et al., 2017). Through these technologies, researchers are describing deeply the different hierarchical levels of plant organization, improving the possibility to predict the behavior of whole plant (phenome). Based on extensive analyses of gene expression (genome and transcriptome) and/or metabolic networks (metabolome), it has been possible to monitor and control cellular responses to genetic perturbations or environmental changes (Fukushima et al., 2009).

However, different constraints can make both the predictability and the controllability difficult from the bottom-up cause-effect approach that underpins the deterministic view of science based on an upward chain of causality (**Figure 1**) (Noble, 2008; Sheth and Thaker, 2014). The first “bottleneck” is how to integrate the massive datasets from molecular high-throughput technologies with the growing high-throughput information on the crop scale, i.e., plant phenomics (Fukushima et al., 2009; Tardieu et al., 2017), which is a typical problem of finding a proper (if it does exist indeed) scaling law (Souza et al., 2016).

The second constraint comes from the common assumption in biochemical models that the system sampled should be in a metabolic steady-state at a given moment, characterized by constant metabolite levels, and that the different metabolic pathways operate in isolation (Toubiana et al., 2013), which is an obvious oversimplification. For instance, at intracellular level, compartmentalization into organelles enables differences in metabolite concentrations acting as a barrier to passive diffusion between organelles and cytoplasm, creating a non-homogeneous cellular metabolic space (Sweetlove and Fernie, 2013; de Souza et al., 2017). Moreover, each metabolic pathway, somehow, is integrated in a dynamical metabolic network (Toubiana et al., 2013), which is challenging for static networks mathematical models that often bypass the network modulation over time. For instance, the stomatal movement depends on a range of environmental and endogenous plant stimuli that affect the internal networks at multiple levels of cellular spatio-temporal organization, generating species-specific responses to combined external stimuli (Merilo et al., 2014). From the modeling of a single guard cell at steady-state, researchers seek to elucidate how these interactions determine the phenotype of plants. However, due to this hierarchy of scales, the interpretation of a large set of data from OMICS tools becomes quite difficult, then it is necessary to develop new methods to allow investigations of dynamic aspects of large scale models (Medeiros et al., 2015).

## OPEN ACCESS

### Edited by:

Paulo Mazzafera,  
Universidade Estadual de Campinas,  
Brazil

### Reviewed by:

Anthony Trewavas,  
University of Edinburgh,  
United Kingdom

### \*Correspondence:

Gustavo M. Souza  
gmsouza@ufpel.edu.br

### Specialty section:

This article was submitted to  
Plant Systems and Synthetic Biology,  
a section of the journal  
Frontiers in Plant Science

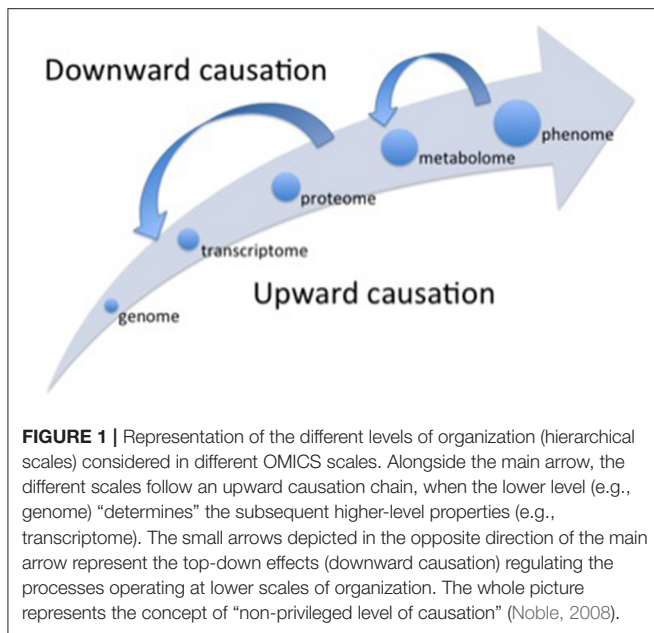
**Received:** 13 September 2017

**Accepted:** 04 December 2017

**Published:** 13 December 2017

### Citation:

do Amaral MN and Souza GM (2017)  
The Challenge to Translate OMICS  
Data to Whole Plant Physiology: The  
Context Matters.  
Front. Plant Sci. 8:2146.  
doi: 10.3389/fpls.2017.02146



The first “bottleneck” refers to the problem of emergent properties at the higher level organizations of the system that are not fully determined by the properties of the lower levels (Souza et al., 2016), for example, changes in the transcriptome or in the proteome do not always result in respective alterations in the metabolome (biochemical phenotype) that exhibits its own dynamics (Ryan and Robards, 2006). Additionally, there is the influence of downward causation processes (Noble, 2008) (Figure 1), when higher levels of organization affect the functioning of the lower levels. For instance, interlocked transcriptional/translational feedback loops are involved in the generation of circadian rhythm in plants, and the functional clock (higher level) controls a wide range of cellular processes such as gene expression (lower level; Fukushima et al., 2009).

The second constraint is related to the different sources of “uncertainty” operating in different levels of plant organization, blurring the predictability from lower levels. For instance, at cellular level, the sources of uncertainty emerge from the spontaneous thermodynamical noise of molecular activity constraining the flux of distributions in metabolic networks (Hoppe et al., 2007), interactions between genes that enable alternative routes for the same phenotype (Kohl et al., 2010), and epigenetic effects changing genes expression (Crisp et al., 2016). Moreover, at the level of whole plant integration, there are many types of long-distance signaling processes (chemical and electrical) overtoping each other and engendering a highly complex informational network that feedback on the regulation of cells metabolism (Choi et al., 2016).

Further, it’s worth to consider that, especially under stressful conditions, the different sources of external “noise” often affect the way that plants respond to environmental changes (Bertolli and Souza, 2013; Prash and Sonnewald, 2015). Environmental fluctuations potentiate the accumulation of conserved cellular signals such as reactive oxygen species (ROS) and the modulation

of intracellular  $\text{Ca}^{2+}$  (Chi et al., 2015; de Souza et al., 2017). Different types of ROS and oxidized molecules produced in different subcellular compartments, together with a spatial and temporal modulation of  $\text{Ca}^{2+}$  elicit different transcriptional responses and, in several cases, the expression of nuclear genes can be altered without altering the total concentration of the signaling molecule in the cell as a whole (Tuteja and Mahajan, 2007; Leister, 2012).

## INTERCELLULAR AND WHOLE PLANT SIGNALING POTENTIATE THE CHALLENGES TO INTEGRATE OMICS INFORMATION ACROSS SCALES

In addition to intracellular complexity, the interaction between neighboring cells plays an important role in the responses to environmental conditions, and the plant metabolism organization as a whole. In plants, plasmodesmata connect the cytoplasts of adjacent cells across the cell wall, allowing intercellular transport and communication to adjacent cells within a tissue or organ, allowing exchange of small molecules, such as ions, sugars, and phytohormones, as well as larger molecules, including proteins, RNA, and viruses (Brunkard et al., 2013). The exchanges of different types of molecules among cells generate, within the same tissue or organ, different gradients of molecules and metabolites, increasing the complexity of physiological processes. In addition, the same type of signal often induces different calcium-dependent responses between two cells of the same type (Gilroy and Trewavas, 2001). Intercellular communication through plasmodesmata plays a crucial role in specifying the fate of cells, as well as in different responses of the same tissue to environmental conditions (Pyott and Molnar, 2015). An example of this can be seen in one of the mechanisms of root development regulation through the short-range cell-to-cell movement of *miR165/6* (Carlsbecker et al., 2010). The expression of mobile *miR165/6* in the endoderm results in a morphogenic gradient, which extends into the xylem layers toward the root center. This generates an opposite *PHABULOSA* (*PHB*) expression gradient (regulated by *miR165/6*), which therefore has a higher concentration in internal xylem tissue. Thus, xylem tissue within the stele is defined, among other factors, by the expression of *PHB*, which is restricted to xylem and procambium by *miR165/6*, specifically expressed in the endoderm.

This intercellular communication through non-autonomous mobile signals adds a further challenge to OMICS approaches, because organs and plant tissues present a great heterogeneity in expression patterns and metabolite profiles, and this information can be lost upon tissue homogenization for downstream analyses.

Besides local communication mechanisms, plants developed long distance signaling processes that enable communication and systemic responses. This type of communication responds to a wide range of environmental stimuli in which the perceived signals are transmitted to the distal organs, inducing systemic responses. Several messengers have been proposed to mediate this systemic communication in plants such as ROS, electrical

signals and  $\text{Ca}^{2+}$ , appearing to be integrated, demonstrating a fast, complex, and finely tuned communication system (Gilroy et al., 2014).

The systemic responses increase system complexity (plant as a whole), and thus increase the uncertainties of the bottom-up predictability models, since physiological changes in specific tissues may have non-local causes. For instance, a local application of high light results in the activation of a ROS wave, allowing an increase in stress tolerance accompanied by the accumulation of photorespiratory amino acids, including Glycine and Serine, in non-stimulated tissues (Suzuki et al., 2013). Examples also demonstrate that  $\text{Ca}^{2+}$  propagation increase in aerial parts of the plant induced by local root treatment with NaCl, showing propagation kinetics differences of leaf-to-leaf (Xiong et al., 2014). The miRNAs also act on systemic responses, such as *miR399* that function as a signaling molecule between the aerial tissues and roots to regulate the uptake of inorganic phosphates (Pi) (Chiou et al., 2006).

Actually, the cell is the result of the properties that emerge from the complex interactions and spatial structures among the thousands of molecules and enzymes of which it is composed. In addition, the environmental context, whether from outside or inside the plant, contributes to shape the way that information is processed by each cell (Gilroy and Trewavas, 2001), and these properties expand at different scales within the plant (Souza et al., 2016). According to Vítolo et al. (2012), the observation of different scales of plant organization, under the same circumstances, can show remarkable differences in the responses to the same stimuli, allowing different interpretations if considering each scale as isolated (Stressed or not stressed? It's the question...). For instance, when plants of soybean were subjected to drought, it was observed, in one hand, significant decreases in gas exchanges (reflecting reduction in plant growth) but, on the other hand, non-significant alterations in chlorophyll fluorescence neither in enzymatic antioxidant activity (Bertolli et al., 2014). Therefore, different scales of organization can show different homeostatic capacities when disturbed, supporting the hypothesis that there is not a privileged level of causation in biological systems (Noble, 2012).

## CONCLUDING REMARKS

The Cartesian method proposed by René Descartes in the seventeenth century set that the first step in order to understand some natural phenomenon is to analyze it, i.e., to decompose the phenomenon in its constitutive parts and to understand them

separately. This first step is based on the mechanistic assumption that the ultimate components of a particular phenomenon “determine” the properties of the phenomenon itself, supporting the raising of the reductionist approach. The second main step in the Cartesian method is the synthesis, i.e., from knowledge gathered of each isolated part build the “whole picture”. Thus, starting from the Galilean scientific revolution (sixteenth century) until the end of twenty and beginning of twenty-first centuries, occidental science was successful to uncover the layers of complexity underlying to the biological organisms, opening the OMICS era with the genome. But the “whole picture” was not clear yet. As exemplified in the previous sections, some problems have challenged the determinism from below, and then the Cartesian synthesis was boosted to explore higher levels of organization, inaugurating the System Biology thinking. The knowledge that has been built on transcriptome, proteome and, specially, the metabolome (Ryan and Robards, 2006) has showed that the higher levels of organization contribute to regulate the lower levels in a downward causation chain (Figure 1), indicating that there is no a privileged level of causation in the organization of biological systems (Noble, 2012).

Therefore, the main message herein is: the context matters. Whatever the scale of observation is taken (from genes to the whole plant), the interpretation of the data shall consider the context in which the particular scale is embedded. Ultimately, in studies that intent to contribute for improvement of crop yield, the plant phenotype (biomass, root, and shoot architecture and/or the crop yield) should have the final word of the meaning of the changes from the lower levels of organization, since one genotype can be translated in many phenotypes when developed under different environmental conditions (Tardieu et al., 2017). Thus, studies tacking into account specific lower levels of organization should maintain their interpretation restricted to those particular levels, avoiding excessive speculative inferences on the higher levels.

## AUTHOR CONTRIBUTIONS

MdA and GS contributed equally to the discussion of the topic.

## ACKNOWLEDGMENTS

The authors thanks to CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) for the fellowships granted to MdA and GS, respectively.

## REFERENCES

- Bertolli, S. C., Mazzafera, P., and Souza, G. M. (2014). Why is it so difficult to identify a single indicator of water stress in plants? a proposal for a multivariate analysis to assess emergent properties. *Plant Biol.* 16, 578–585. doi: 10.1111/plb.12088
- Bertolli, S. C., and Souza, G. M. (2013). The level of environmental noise affects the physiological performance of Glycine max under water deficit. *Theor. Exp. Plant Physiol.* 25, 36–45. doi: 10.1590/S2197-00252013000100005
- Brunkard, J. O., Runkel, A. M., and Zambryski, P. C. (2013). Plasmodesmata dynamics are coordinated by intracellular signaling pathways. *Curr. Opin. Plant Biol.* 16, 614–620. doi: 10.1016/j.pbi.2013.07.007
- 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
- Chi, W., Feng, P., Ma, J., and Zhang, L. (2015). Metabolites and chloroplast retrograde signaling. *Curr. Opin. Plant Biol.* 25, 32–38. doi: 10.1016/j.pbi.2015.04.006

- Chiou, T. J., Aung, K., Lin, S. I., Wu, C. C., Chiang, S. F., and Su, C. L. (2006). Regulation of phosphate homeostasis by microRNA in Arabidopsis. *Plant Cell* 18, 412–421. doi: 10.1105/tpc.105.038943
- Choi, W. G., Hilleary, R., Swanson, S. J., Kim, S. H., and Gilroy, S. (2016). Rapid, long distance electrical and calcium signaling in plants. *Annu. Rev. Plant Biol.* 67, 287–307. doi: 10.1146/annurev-arplant-043015-112130
- Crisp, P. A., Ganguly, D., Eichten, S. R., Borevitz, J. O., and Pogson, B. J. (2016). Reconsidering plant memory: intersections between stress recovery, RNA turnover, and epigenetics. *Sci. Adv.* 2:e1501340. doi: 10.1126/sciadv.1501340
- de Souza, A., Wang, J. Z., and Dehesh, K. (2017). Retrograde signals: integrators of interorganellar communication and orchestrators of plant development. *Annu. Rev. Plant Biol.* 68, 58–108. doi: 10.1146/annurev-arplant-042916-041007
- Edwards, D., and Batley, J. (2004). Plant bioinformatics: from genome to phenome. *Trends Biotechnol.* 22, 232–237. doi: 10.1016/j.tibtech.2004.03.002
- Fukushima, A., Kusano, M., Nakamichi, N., Kobayashi, M., Hayashi, N., Sakakibara, H., et al. (2009). Impact of clock-associated Arabidopsis pseudo-response regulators in metabolic coordination. *Proc. Natl. Acad. Sci. U.S.A.* 106, 7251–7256. doi: 10.1073/pnas.0900952106
- Gilroy, S., Suzuki, N., Miller, G., Choi, W.-G., Toyota, M., Deviredy, A. R., et al. (2014). A tidal wave of signals: calcium and ROS at the forefront of rapid systemic signaling. *Trends Plant Sci.* 19, 623–630. doi: 10.1016/j.tplants.2014.06.013
- Gilroy, S., and Trewavas, A. (2001). Signal processing and transduction in plant cells: the end of the beginning? *Nat. Rev. Mol. Cell Biol.* 2, 307–314. doi: 10.1038/35067109
- Hoppe, A., Hoffmann, S., and Holzthütter, H.-G. (2007). Including metabolite concentrations into flux balance analysis: thermodynamic realizability as a constraint on flux distributions in metabolic networks. *BMC Syst. Biol.* 1:23. doi: 10.1186/1752-0509-1-23
- Kohl, P., Crampin, E. J., Quinn, T. A., and Noble, D. (2010). Systems biology: an approach. *Clin. Pharmacol. Ther.* 88, 25–33. doi: 10.1038/clpt.2010.92
- Leister, D. (2012). Retrograde signaling in plants: from simple to complex scenarios. *Front. Plant Sci.* 3:135. doi: 10.3389/fpls.2012.00135
- Medeiros, D. B., Daloso, D. M., Fernie, A. R., Nikoloski, Z., and Araújo, W. L. (2015). Utilizing systems biology to unravel stomatal function and the hierarchies underpinning its control. *Plant Cell Environ.* 38, 1457–1470. doi: 10.1111/pce.12517
- Merilo, E., Jöesaar, I., Brosché, M., and Kollist, H. (2014). To open or to close: species-specific stomatal responses to simultaneously applied opposing environmental factors. *New Phytol.* 202, 499–508. doi: 10.1111/nph.12667
- Mochida, K., and Shinozaki, K. (2011). Advances in omics and bioinformatics tools for systems analyses of plant functions. *Plant Cell Physiol.* 52, 2017–2038. doi: 10.1093/pcp/pcr153
- Noble, D. (2008). Claude Bernard, the first systems biologist, and the future of physiology. *Exp. Physiol.* 93, 16–26. doi: 10.1113/expphysiol.2007.038695
- Noble, D. (2012). A theory of biological relativity: no privileged level of causation. *Interface Focus* 2, 55–64. doi: 10.1098/rsfs.2011.0067
- Prasch, C. M., and Sonnewald, U. (2015). Signaling events in plants: stress factors in combination change the picture. *Environ. Exp. Bot.* 114, 4–14. doi: 10.1016/j.envexpbot.2014.06.020
- Pyott, D. E., and Molnar, A. (2015). Going mobile: non-cell-autonomous small RNAs shape the genetic landscape of plants. *Plant Biotechnol. J.* 13, 306–318. doi: 10.1111/pbi.12353
- Ryan, D., and Robards, K. (2006). Metabolomics: the greatest omics of them all? *Anal. Chem.* 78, 7954–7958. doi: 10.1021/ac0614341
- Sheth, B. P., and Thaker, V. S. (2014). Plant systems biology: insights, advances and challenges. *Planta* 240, 33–54. doi: 10.1007/s00425-014-2059-5
- Souza, G. M., Bertolli, S. C., and Lüttge, U. (2016). “Hierarchy and information in a system approach to plant biology: explaining the irreducibility in plant ecophysiology” in *Progress in Botany*, eds U. Lüttge (Berlin: Springer Science+Business Media), 167–186.
- Suzuki, N., Miller, G., Salazar, C., Mondal, H. A., Shulaev, E., Cortes, D. F., et al. (2013). Temporal-spatial interaction between reactive oxygen species and abscisic acid regulates rapid systemic acclimation in plants. *Plant Cell* 25, 3553–3569. doi: 10.1105/tpc.113.114595
- Sweetlove, L. J., and Fernie, A. R. (2013). The spatial organization of metabolism within the plant cell. *Annu. Rev. Plant Biol.* 64, 723–746. doi: 10.1146/annurev-arplant-050312-120233
- Tardieu, F., Varshney, R. K., and Tuberosa, R. (2017). Improving crop performance under drought – cross-fertilization of disciplines. *J. Exp. Bot.* 68, 1393–1398. doi: 10.1093/jxb/erx042
- Toubiana, D., Fernie, A. R., Nikoloski, Z., and Fait, A. (2013). Network analysis: tackling complex data to study plant metabolism. *Trends Biotechnol.* 31, 29–36. doi: 10.1016/j.tibtech.2012.10.011
- Tuteja, N., and Mahajan, S. (2007). Calcium signaling network in plants: an overview. *Plant Signal. Behav.* 2, 79–85. doi: 10.4161/psb.2.2.4176
- Vítole, H. F., Souza, G. M., and Silveira, J. A. G. (2012). Cross-scale multivariate analysis of physiological responses to high temperature in two tropical crops with C-3 and C-4 metabolism. *Environ. Exp. Bot.* 80, 54–62. doi: 10.1016/j.envexpbot.2012.02.002
- Xiong, T. C., Ronzier, E., Sanchez, F., Corratgé-Faillie, C., Mazars, C., and Thibaud, J.-B. (2014). Imaging long distance propagating calcium signals in intact plant leaves with the BRET-based GFP-aequorin reporter. *Front. Plant Sci.* 5:43. doi: 10.3389/fpls.2014.00043

**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 © 2017 do Amaral and Souza. 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) or licensor 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.





# Inference of Transcription Regulatory Network in Low Phytic Acid Soybean Seeds

Neelam Redekar<sup>1†</sup>, Guillaume Pilot<sup>2</sup>, Victor Raboy<sup>3</sup>, Song Li<sup>1\*</sup> and M. A. Saghai Maroof<sup>1\*</sup>

<sup>1</sup> Department of Crop and Soil Environmental Sciences, Virginia Tech, Blacksburg, VA, United States, <sup>2</sup> Department of Plant Pathology, Physiology, and Weed Science, Virginia Tech, Blacksburg, VA, United States, <sup>3</sup> National Small Grains Germplasm Research Center, Agricultural Research Service (USDA), Aberdeen, ID, United States

## OPEN ACCESS

### Edited by:

Diego Mauricio Riaño-Pachón,  
Institute of Chemistry, University of  
São Paulo, Brazil

### Reviewed by:

Jedreż Jakub Szymanski,  
Weizmann Institute of Science, Israel  
Sudip Kundu,  
University of Calcutta, India  
Maria Katherine Mejia Guerra,  
Cornell University, United States

### \*Correspondence:

Song Li  
songli@vt.edu  
M. A. Saghai Maroof  
smaroof@vt.edu

### †Present Address:

Neelam Redekar,  
Department of Crop and Soil Science,  
Oregon State University, Corvallis, OR,  
United States

### Specialty section:

This article was submitted to  
Plant Systems and Synthetic Biology,  
a section of the journal  
Frontiers in Plant Science

**Received:** 24 July 2017

**Accepted:** 14 November 2017

**Published:** 30 November 2017

### Citation:

Redekar N, Pilot G, Raboy V, Li S and  
Saghai Maroof MA (2017) Inference of  
Transcription Regulatory Network in  
Low Phytic Acid Soybean Seeds.  
Front. Plant Sci. 8:2029.  
doi: 10.3389/fpls.2017.02029

A dominant loss of function mutation in *myo*-inositol phosphate synthase (*MIPS*) gene and recessive loss of function mutations in two multidrug resistant protein type-ABC transporter genes not only reduce the seed phytic acid levels in soybean, but also affect the pathways associated with seed development, ultimately resulting in low emergence. To understand the regulatory mechanisms and identify key genes that intervene in the seed development process in low phytic acid crops, we performed computational inference of gene regulatory networks in low and normal phytic acid soybeans using a time course transcriptomic data and multiple network inference algorithms. We identified a set of putative candidate transcription factors and their regulatory interactions with genes that have functions in *myo*-inositol biosynthesis, auxin-ABA signaling, and seed dormancy. We evaluated the performance of our unsupervised network inference method by comparing the predicted regulatory network with published regulatory interactions in *Arabidopsis*. Some contrasting regulatory interactions were observed in low phytic acid mutants compared to non-mutant lines. These findings provide important hypotheses on expression regulation of *myo*-inositol metabolism and phytohormone signaling in developing low phytic acid soybeans. The computational pipeline used for unsupervised network learning in this study is provided as open source software and is freely available at <https://lilabatvt.github.io/LPANetwork/>.

**Keywords:** phytic acid, soybean seed development, *myo*-inositol metabolism, unsupervised machine learning, gene regulatory network

## INTRODUCTION

Seed development is a complex metabolic process, which involves both synthesis and breakdown of macromolecules for growth and maintenance of the embryo (Weber et al., 2005; Le et al., 2007). During seed development, glucose-6-phosphate is converted to *myo*-inositol, an intracellular signaling molecule, which is phosphorylated several times to form phytic acid (Raboy, 1997). Seeds with reduced phytic acid content are commercially more valuable because consumption of low phytic acid seeds by monogastric animals alleviates mineral deficiency and reduces phosphorus pollution from animal waste (Raboy, 2007). Mutations that block the phytic acid biosynthesis

**Abbreviations:** *lpa*, Low phytic acid; MIPS, *myo*-inositol phosphate synthase; MRP, Multi-drug resistance protein; DEG, Differentially expressed gene; RFO, Raffinose family oligosaccharide; ABA, Absciscic acid.

pathway have been shown to alter the seed metabolite levels in soybean, rice, maize, and other plant species (Wilcox et al., 2000; Shi et al., 2003, 2005; Stevenson-Paulik et al., 2005; Raboy, 2007; Glover, 2011; Jervis et al., 2015). For example, a mutation in *myo*-inositol phosphate synthase (*MIPS*) gene results in reduced phytic acid, stachyose, raffinose, and elevated sucrose, and low seed emergence in soybean (Hitz et al., 2002; Saghai Maroof and Buss, 2008). Other non-biosynthetic pathway genes such as multi-drug resistance protein (*MRP*) genes encoding ATP-binding cassette transporters that are believed to be involved in the transport of phytic acid to storage vacuoles, are also known to regulate phytic acid levels and affect seed emergence (Shi et al., 2007; Nagy et al., 2009; Saghai Maroof et al., 2009; Xu et al., 2009; Jervis et al., 2015).

Transcriptome analysis is a valuable tool for the characterization of the regulatory networks that mediate this complex interaction. The expressions of genes involved in the metabolic activities in seeds are tightly regulated by the synergistic action of many transcription factors and other regulatory genes (Weber et al., 2005; Le et al., 2007). Two independent studies, one with barley low phytic acid (*lpa*) mutant, and another with soybean *mips1/mrp-l/mrp-n* (*'3mlpa'*) triple mutant have reported the effect of *lpa* mutations on the transcriptomic profiles of developing seeds (Bowen et al., 2007; Redekar et al., 2015). The differential expression of transcription factor genes such as WRKY and CAMTA (Calmodulin-binding Transcription Activator), was linked to phytic acid biosynthesis pathway, suggesting a complex regulatory mechanism (Redekar et al., 2015). Association of WRKY transcription factors and  $Ca^{2+}$  binding activity with inositol metabolism was also confirmed by another independent study with maize low phytic acid breeding line Qi319 (Zhang et al., 2016). Zhang et al. (2016) also identified ABC transporter gene candidates associated with low phytic acid phenotype in maize using co-regulatory network. In this article, we focus on the discovery of transcription regulatory networks to further investigate the inositol metabolism in soybean.

One type of the widely applied method of network inference is the use of Pearson Correlation Coefficient or related methods for data analysis (Langfelder and Horvath, 2008; Bassel et al., 2011; Li et al., 2012). Although such correlation analyses can cluster genes with similar functions, the resultant networks do not predict the direction of gene regulation. Many other approaches, such as mutual information (Faith et al., 2007), partial correlation (Faith et al., 2007), random forest (Huynh-Thu et al., 2010), and least angle regression (LARS) (Haury et al., 2012), have been developed to perform inference of directed gene regulatory networks. For well-characterized model organisms such as *Arabidopsis*, known interactions from ChIP-chips or ChIP-seq experiments can be used as prior knowledge in supervised machine learning approaches (Maetschke et al., 2014; Ni et al., 2016). However, for biological systems where little prior information is available, such as in soybeans, unsupervised methods have to be used for network inference. In particular, three methods including co-expression analysis (Bassel et al., 2011), decision trees (Zhu et al., 2013) and mutual information (Gonzalez-Morales et al., 2016) have been successfully applied to

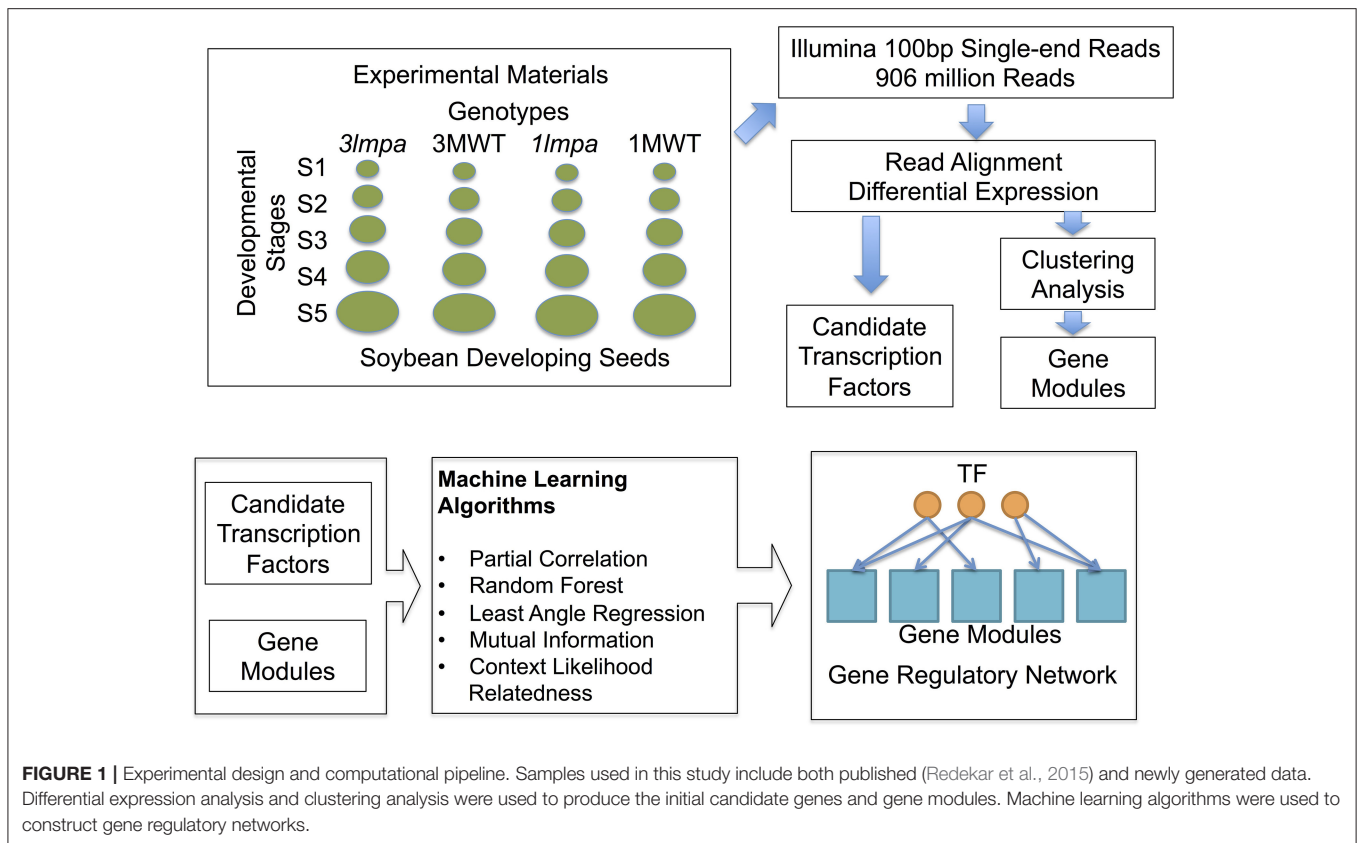
identify functional networks in *Arabidopsis* and soybean seeds. With numerous inference methods available, it has been found that congregating the prediction results from multiple methods (so-called “community-based method”) improves the prediction accuracy as compared to any individual method (Marbach et al., 2012).

In this study, we performed computational inference of gene regulatory networks in low phytic acid mutants and the corresponding non-mutant soybean seeds from time course transcriptomic data. In addition to previously published RNA-seq data (Redekar et al., 2015), we generated new RNA-seq data of developing seeds using a pair of soybean isogenic lines, one carrying the *mips1* mutation (*'1mlpa'*) and the other the corresponding wild type allele. We implemented a computational pipeline for unsupervised gene regulatory network inference using five different methods: ARACNE (Margolin et al., 2006), Random Forest (Huynh-Thu et al., 2010), LARS (Haury et al., 2012), partial correlation (Schafer and Strimmer, 2005b), and context likelihood relatedness (CLR) (Faith et al., 2007). To improve computational efficiency and interpretability of the inferred network, we adopted the widely used module network approach by which genes were grouped into co-expression modules and inference of gene regulation was performed between transcription factors and gene modules (Segal et al., 2003). We found that many gene modules included genes with meaningful biological functions and some gene modules showed genotype-specific expression patterns. We identified several transcription factors that were differentially expressed between developmental stages and some of the inferred regulatory interactions were specifically found in mutants or non-mutants. Genes involved in phytic acid metabolism and related metabolic processes were found in multiple modules and were predicted to be regulated by different transcription factors. For validation, the predicted interactions were compared with known regulatory interactions observed in the model plant species, *Arabidopsis thaliana*. These findings provide important hypotheses on expression regulation of *myo*-inositol metabolism, and phytohormone signaling in developing *lpa* soybeans. The computational method for inferring regulatory networks is freely available at <https://lilabvt.github.io/LPANetwork/>. This method can be used to perform network inference using time series data from soybean or any other crop species.

## MATERIALS AND METHODS

### Genetic Materials

Four soybean experimental lines designated as: (i) *3mlpa*, (ii) 3MWT, (iii) *1mlpa*, and (iv) 1MWT were used in this study (Figure 1). The *lpa* mutant line, *'3mlpa'*, carrying three mutations *mips1/mrp-l/mrp-n*, and its non-mutant sibling line with normal phytic acid, '3MWT', were derived from crossing of 'CX-1834' (*lpa* line with two *mrp-l/mrp-n* mutations on soybean chromosomes 19 and 3, respectively) with 'V99-5089' (*lpa* line with single *mips1* mutation) (Saghai Maroof et al., 2009). The low phytic acid causing mutations in the parental lines have been mapped to genes Glyma.11G238800 (*MIPS1*),



Glyma.19g169000 (MRP-L), and Glyma.03g167800 (MRP-N) (Saghai Maroof et al., 2009). Another *lpa* line, '1mlpa', carrying a single *mips1* mutation on soybean chromosome 11, and its isogenic sibling line with normal phytic acid, '1MWT', were derived from crossing of 'Essex' (a normal phytic acid line with no mutations) with V99-5089 (Saghai Maroof and Buss, 2008; Glover, 2011).

## Plant Growth and Tissue Sampling

Four seeds from each of experimental lines—3mlpa, 3MWT, 1mlpa, and 1MWT, were planted in each of 12 pots containing Metro-Mix<sup>®</sup> 360 (Sun Gro) media topped with GardenPro ULTRALITE soil (Redekar et al., 2015). These plants (48/line) were grown in growth chambers with 14 /10 h photoperiod, 24°C/16°C temperature, 300–400  $\mu$ E light and 50–60% relative humidity. Developing seeds were sampled in triplicates for each experimental line based on seed lengths corresponding to 2–4 mm (S1), 4–6 mm (S2), 6–8 mm (S3), 8–10 mm (S4), and 10–12 mm (S5), respectively. Samples were flash frozen using liquid nitrogen and stored at  $-70^{\circ}\text{C}$ . High-quality total RNA (RIN 9–10) was extracted from frozen samples using RNeasy Plant Mini Kit, with on-column DNase digestion (QIAGEN). Total of 60 mRNA libraries were prepared from total RNA samples and sequenced as 100SE using HiSeq2000 at the Genome Quebec Innovation Center, Canada.

## Sequence Data Processing and Differential Gene Expression

Reads were aligned to the latest soybean reference genome ('Williams 82' Wm82.a2.v1, downloaded from Phytozome<sup>1</sup> with STAR (version 2.4.2) and number of reads mapped to each gene was counted using featureCount (version 1.4.6). Differential gene expression was analyzed using DESeq2 (version 1.8.2) in R (version 3.2.4). Four genotypes, 3mlpa, 3MWT, 1mlpa, and 1MWT, were analyzed in this data. For each pair of mutant and corresponding non-mutant, stage-wise comparison was performed to identify differentially expressed genes for each stage (Supplementary Figure 1A) (Redekar et al., 2015). For each genotype, between-stage comparisons were performed to identify differentially expressed genes between adjacent developmental stages (Supplementary Figure 1B). These analyses were performed using DESeq2 with default parameters. Between-stage comparisons and stage-wise comparisons address different type of biological question. Between-stage comparisons find genes that change between stages, but do not directly identify genes that are affected by mutations. Stage-wise comparison, on the other hand, directly finds genes that change between mutant and non-mutant lines, but does not find genes with interactions. Differentially expressed genes were the genes with FDR adjusted  $p < 0.01$  and  $\log_2$  fold change  $> 1$ . Differentially expressed genes and their  $\log_2$  fold changes are provided as Supplementary Tables

<sup>1</sup><https://phytozome.jgi.doe.gov/>

1–3. RNA-Seq data used in this study have been deposited into the NCBI Gene Expression Omnibus (GEO) repository under accession number GSE101692.

## Inference of Gene Regulatory Networks Expression Clustering, Gene Ontology, and Gene Function Analysis

Gene expression levels for each gene were normalized using DESeq2 and summarized as FPKM (Fragments Per Kilo-base pair per Million reads) values. The gene expression levels (FPKM values) were averaged across replicates and only differentially expressed genes were used in the clustering analysis. K-means clustering (Sherlock, 2000) was performed using R packages, and the number of clusters (K) was determined using the minimum Bayesian Information Criteria (BIC) method (Ramsey et al., 2008). In brief, K was set to be an integer number from 20 to 100 with an incremental step size of 5. For each K-value, k-means clustering was performed and BIC statistics were computed. The minimum BIC was achieved with  $K = 60$  (Supplementary Figure 2). Gene Ontology (GO) annotation of all soybean genes was downloaded from Soybase<sup>2</sup> GO enrichment analyses were performed for each gene module. Significantly enriched GO categories were selected using Fisher's exact test with FDR <0.05 (Supplementary Table 4). Transcription factor annotation was downloaded from plant TFDB (Jin et al., 2015, 2017). Metabolic pathway genes were downloaded from the SoyCyc 7.0<sup>3</sup> database from the Plant Metabolic Network<sup>4</sup> website.

## Network Inference Methods

To infer regulatory networks, we adopted the methods of module networks (Segal et al., 2003). First, genes were grouped into modules using the k-means clustering method. Second, differentially expressed transcription factors were used as putative regulators for network inference. In our data, we found 60 clusters (gene modules) and 1245 transcription factors that were differentially expressed in at least one comparison. The mean expression profile for each of the 60 modules was computed and the expression levels of 1245 transcription factors were included to construct an expression matrix with 1305 rows (genes) and 20 columns (five developmental stages for four experimental lines). Five distinct network inference algorithms: ARACNE (Margolin et al., 2006), Random Forest (Huynh-Thu et al., 2010), LARS (Haury et al., 2012), partial correlation (Schafer and Strimmer, 2005b), and CLR (Faith et al., 2007), were applied to this expression matrix to infer putative regulatory interactions between each transcription factor and gene modules. These methods were chosen because they represent a diverse set of computational methods for gene network inference. These methods were selected also because they were ranked as top performers in a recently published benchmark of network inference methods (DREAM challenge) (Marbach et al., 2012). Details of each method, statistical

analysis of network and network validation are provided as supplementary text.

## RESULTS

### Summary of Differential Gene Expression Analysis

Transcriptome sequencing data from five developing seed stages of four soybean lines (*3mlpa*, 3MWT, *1mlpa*, and 1MWT) were analyzed (Figure 1). Stage-wise comparisons were performed for mutants (*3mlpa*, *1mlpa*) and their corresponding non-mutants (3MWT, 1MWT), to determine the number of genes affected by the mutation at each stage (Figure 2). For stage-wise comparisons, when *1mlpa* was compared with 1MWT, we found fewer than 250 differentially expressed genes in all time points (Supplementary Figure 2). However, we found more than 4000 differentially expressed genes between *3mlpa* and 3MWT (Supplementary Figure 2). It is expected to have higher number of differentially expressed genes for comparison between recombinant inbred lines (*3mlpa* vs. 3MWT) as opposed to that between near-isogenic lines (*1mlpa* vs. 1MWT). Few genes are differentially expressed in all five stages when comparing *3mlpa* vs. 3MWT (Figure 2B). These results suggest that genes affected by mutations are unique at each developmental stage. Between-stages comparisons were performed for each genotype separately (Figure 1, Supplementary Figure 1). Results of the stage-wise and between-stages differential expression analyses are provided as Supplementary Tables 1, 2, respectively. For between-stages comparisons (Supplementary Figure 1B), we found that hundreds of genes are differentially expressed when comparing adjacent developmental stages (Supplementary Figure 2). However, there are few genes differentially expressed across all stages. For example, in *3mlpa* mutant, only 2 genes are differentially expressed between any two adjacent stages (Figure 2A). We found 1643 genes that are differentially expressed between developmental stages and found in all four genotypes in this study (Figure 2C), suggesting these genes are a core set of genes that change expression between developmental stages and are not affected by the genotypes. We determined the number of differentially expressed genes in each of the comparisons (Supplementary Figure 2). Interestingly, the highest numbers of differentially expressed genes were found in two comparisons: between stage 1 and stage 2 for 1MWT and for 3MWT (3MWT S2 vs. S1 and 1MWT S2 vs. S1) suggesting that non-mutant plants have a high number of differentially expressed genes in the early stages of seed development than mutants.

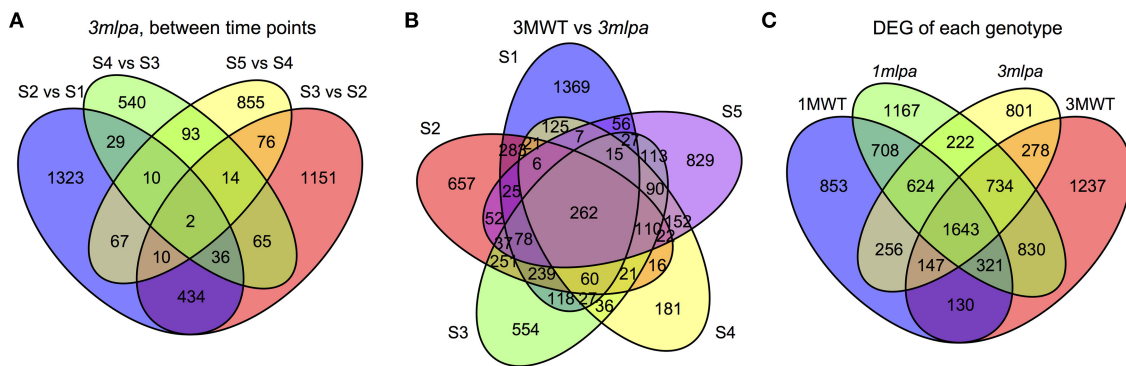
The two sets of low-phytic acid causing mutations (*mips1* and *mrp-l/mrp-n*) interrupt the phytic acid biosynthesis and transmembrane transporter activity, ultimately reducing the seed emergence potential in our mutant lines. We were, therefore, interested in studying the behavior of genes associated with phytic acid metabolism, abscisic acid (ABA) and auxin signaling and metabolism, and transmembrane transport. The log<sub>2</sub> fold change for significantly differentially expressed genes that belonged to this category is summarized in Supplementary Table 3.

<sup>2</sup><https://soybase.org/>

<sup>3</sup><http://www.plantcyc.org/databases/soycyc/7.0>

<sup>4</sup><http://www.plantcyc.org/>





**FIGURE 2 |** Venn diagrams. S1, S2, S3, S4, and S5 represent five developmental stages. **(A)** Number of genes differentially expressed between consecutive stages in *3mlpa* seeds. **(B)** Number of genes differentially expressed for each stage when comparing *3mlpa* to 3MWT seeds. **(C)** Number of genes differentially expressed for different genotypes for between stage comparisons.

## Co-expression Modules Represent Distinct Functional Categories

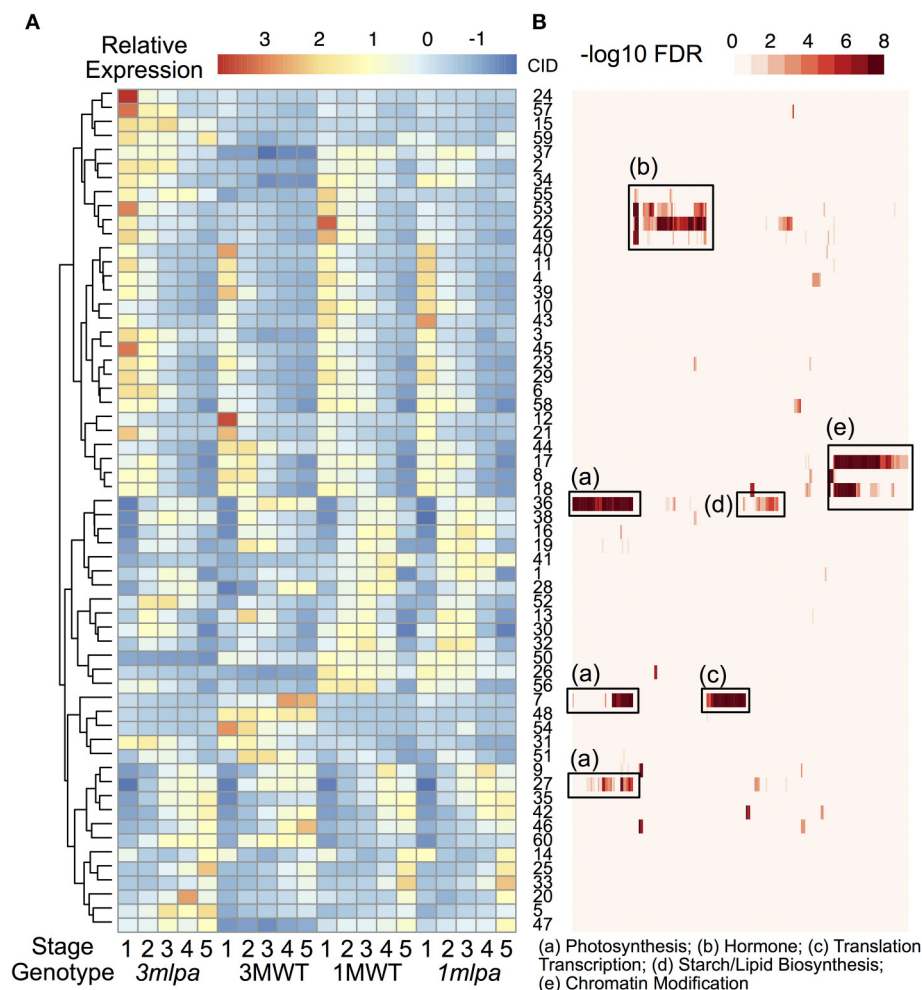
We identified 12998 genes that are differentially expressed in at least one of the comparisons and these genes were used for K-means clustering analysis to identify co-expressed gene modules (Figures 1, 3). We found that the optimal number of clusters (modules) is 60 based on BIC (Supplementary Figure 3). Average expression levels of all genes in each module were used to generate a heat map (Figure 3A) and GO enrichment analyses were performed for each of the modules (Figure 3B, Supplementary Table 4). The expression modules can be approximately classified into three main patterns. In pattern 1, there are 29 modules highly expressed at the early stage of seed development as shown in upper half of the heat map (Figure 3A, cluster 24 to cluster 18). In pattern 2, there are 11 modules highly expressed in the later stage of seed development as shown by the lower portion of the heat map (Figure 3A, modules 9 to 47, except for module 14, which shows high expression at both first and last developmental stages). In pattern 3, 12 modules showed high expression in the middle of the developmental stages but low expression levels in the early and late developmental stages. Genotype-specific expression patterns were also found by clustering analysis. For example, modules 24, 57, and 15 are highly expressed in *3mlpa* at the first time point, whereas the expression levels are not high in the other three genotypes. The modules identified in near isogenic lines (*1mlpa* and 1MWT) showed highly similar expression patterns than those identified in recombinant inbred lines (*3mlpa* and 3MWT).

GO enrichment analyses showed that many gene modules are enriched with genes in specific functional categories (Supplementary Table 4). For example, genes in modules 22, 49, 53, and 55 are highly expressed during the early stages of seed development and are enriched with genes with functions in hormone signaling and responses (Figure 3B, box b). Of these, several genes showed increased expression in *3mlpa* (S1) while decreased in *1mlpa* (S1) when compared with respective non-mutant lines at stage 1 (Table 1). We also

found that genes in module 42 are highly expressed at the last stage of seed development and this module is enriched with genes functioning in seed dormancy, seed germination and lipid storage (Figure 3B). Genes in module 36 are highly expressed in the middle stages of seed development and are enriched with genes in starch and lipid biosynthesis (Figure 3B, box d). The functional enrichment of genes in these modules indicates that our clustering analysis can find genes representing biological functions that are known to be active at different stages of seed development. We also found that some modules have genotype-specific expression patterns. For example, module 7 is highly expressed only in 3MWT, which is enriched with genes functioning in photosynthesis, translations, and transcription (Figure 3B, boxes a,c). This result shows some gene modules have genotype-specific expression pattern and are enriched with specific functional genes.

## Regulatory Network Interactions

Five different network inference algorithms were used to infer putative regulatory interactions between regulators and their targets (Figures 1, 4). Fifty-four interactions between 54 transcription factors and 32 modules were predicted by all five algorithms (Supplementary Table 5). Some modules were predicted to be regulated by more than one transcription factor and no transcription factor was predicted to regulate two modules. The identified interactions represent highly stringent predictions and are a very conservative estimation of all possible interactions since only 0.06% of all 74,700 possible interactions were found to be significant by all five computational methods. Four hundred six interactions between 348 transcription factors and 60 modules were supported by four or more methods (Supplementary Figure 4, Supplementary Table 6), representing a larger number for predicted regulatory interactions. The network figure includes both directed and undirected edges (Figure 4). Each directed edge connects a transcription factor with its targeted gene module. Such edge represents predicted regulatory interactions. Each undirected



**FIGURE 3 |** Clustering and gene function analysis. **(A)** Gene expression clusters. Color indicates normalized expression levels. Genes were clustered based on the K-means clustering algorithm. Hierarchical clustering was performed on the rows of the average expression levels for each cluster. Numbers on the right of the heatmap represent cluster id (CID, 1 to 60). **(B)** GO functional enrichment analysis. Rows of this heatmap is organized as the same order as the expression clusters. Each column represents a GO category. Color represents  $-\log_{10}$  FDR from enrichment analysis. Some enriched categories were summarized and annotated under the heat map. The complete set of enriched GO categories is provided as supplementary information.

edge connects a transcription factor to the module this factor belongs to. These undirected edges reflect the fact that each transcription factor is also co-expressed with other genes in the genome and can be assigned to a specific gene module. The regulatory interactions (directed edges) are further classified based on the differential expression pattern of the regulatory TFs. We found 10 TF-module interactions (black arrows) in which the TFs are differentially expressed between stage comparisons (Supplementary Figure 1B) in both mutants (*3mlpa* and *1mlpa*) and both non-mutants (3MWT and 1MWT). These interactions are not affected by the mutations or genetic backgrounds. We found 10 interactions (Figure 4, green arrows) in which the TFs are differentially expressed between stage comparisons for non-mutants but not in mutants. These interactions are potentially lost due to the mutations. We also found nine interactions (blue arrows) that are not present in the non-mutants, but

are present in either one of the mutants, suggesting that these interactions are gained in the mutants. Finally, we found 14 interactions (red arrows) that are not present in either the mutants or the non-mutants, but the TFs are differentially expressed when comparing *3mlpa* to 3MWT at one or more developmental stages. These interactions are altered in *3mlpa*/3MWT and do not change the trajectory of gene expression between stages but affect gene expression within specific stages. We found that many TFs are also connected to the target modules by undirected edges, indicating that these genes are co-expressed with their target genes. We also found several cases where a TF does not regulate its own module but is regulating other modules, suggesting our method can find non-linear interactions between TF and target modules. All the predicted regulatory interactions are provided as a (Supplementary Tables 5, 6).

**TABLE 1** | Differentially expressed genes from co-expression modules in early stages of seed development.

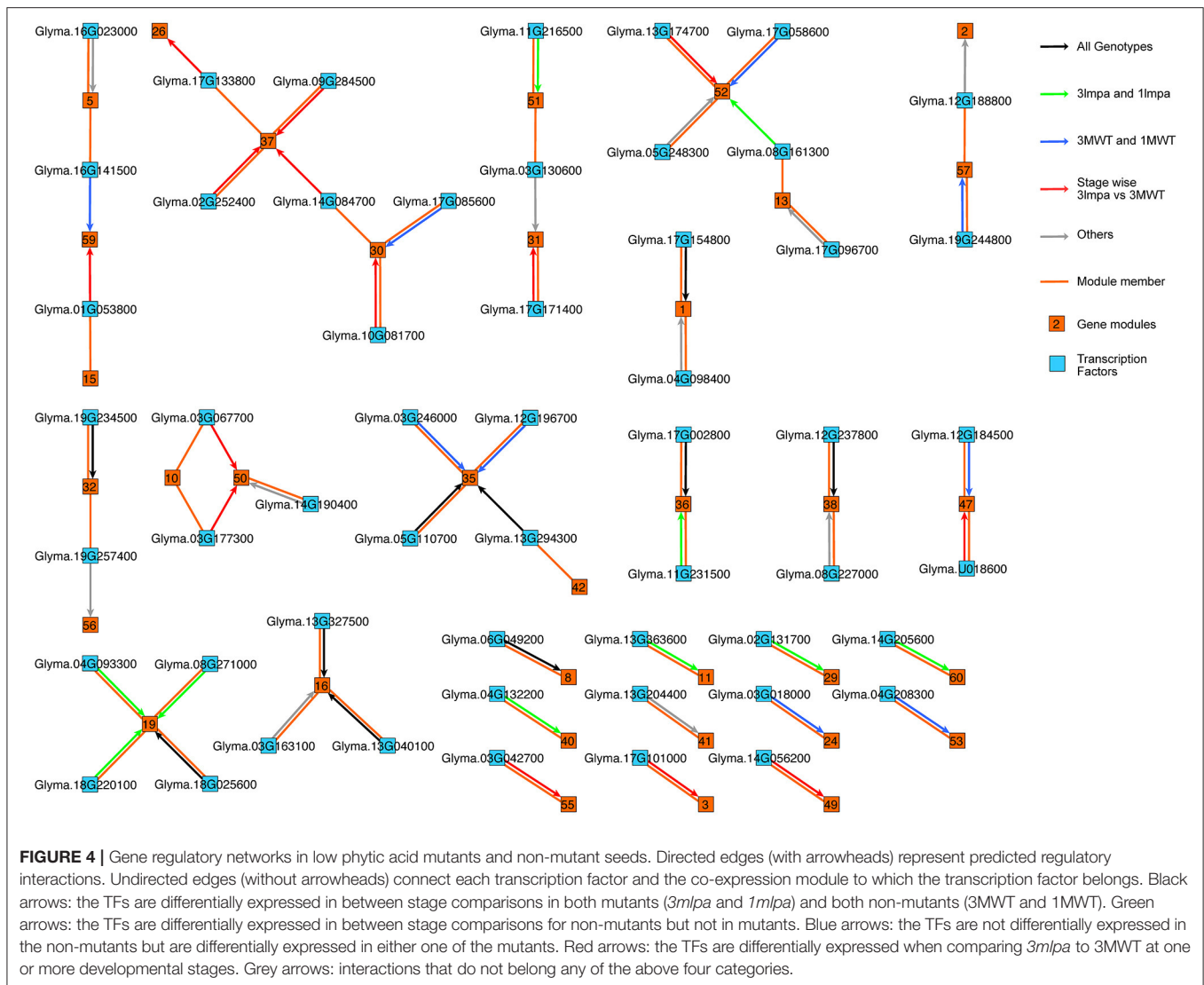
Cluster ID	GO Term	Gene Name	Log2 fold change		Arabidopsis homolog	Gene Symbol	Description
			3mlpa vs. 3MWT-S1	1mlpa vs. 1MWT-S1			
22	GO:0048838: release of seed from dormancy	Glyma.01G153300	3.13805	−2.84401	AT4G19230	<i>CYP707A1</i>	Cytochrome P450, family 707, subfamily A, polypeptide 1
		Glyma.12G073000	2.67451	−2.55671	AT3G45640	<i>MPK3</i>	Mitogen-activated protein kinase 3
	GO:0009738: abscisic acid-activated signaling pathway	Glyma.U021800	2.5946	−2.50453	AT3G45640	<i>MPK3</i>	Mitogen-activated protein kinase 3
		Glyma.07G023300	2.96832	−2.24850	AT1G80840	<i>WRKY40</i>	WRKY DNA-binding protein 40
		Glyma.10G082400	1.40473	−1.55483	AT3G11820	<i>SYP121, SYR1, PEN1</i>	Syntaxin of plants 121
		Glyma.08G271900	2.04156	−1.54559	AT1G32640	<i>RD22BP1, JAI1, JIN1, MYC2, ZBF1</i>	Basic helix-loop-helix (bHLH) DNA-binding family protein
		Glyma.03G138000	0	−1.46586	AT3G57530	<i>CPK32, CDPK32</i>	Calcium-dependent protein kinase 32
		Glyma.05G194100	0	−1.41822	AT3G52430	<i>PAD4</i>	Alpha/beta-Hydrolases superfamily protein
		Glyma.17G126400	1.18340	−1.31950	AT3G48360	<i>BT2</i>	BTB and TAZ domain protein 2
42	GO:0009845: seed germination	Glyma.17G086400	1.56707	0	AT3G01570	-	Oleosin family protein
49	GO:0009738: abscisic acid-activated signaling pathway	Glyma.15G078600	0	−1.84111	AT1G28480	<i>GRX480, roxy19</i>	Thioredoxin superfamily protein
		Glyma.14G066400	0	−1.10638	AT3G11410	<i>PP2CA, AHG3</i>	Protein phosphatase 2CA
53	GO:0009737: response to abscisic acid	Glyma.13G119500	1.60968	−2.02755	AT3G17510	<i>CIPK1, SnRK3.16</i>	CBL-interacting protein kinase 1
		Glyma.12G186800	2.75115	−1.95635	AT2G27310		F-box family protein
	GO:0009738: abscisic acid-activated signaling pathway	Glyma.12G186800	2.75115	−1.95635	AT2G27310		F-box family protein
	GO:0009733: response to auxin	Glyma.13G094900	4.14408	−2.26796	AT5G57560	<i>TCH4, XTH22</i>	Xyloglucan endotransglucosylase/ hydrolase family protein

Validation of Predicted Regulatory Networks

To validate the results from the computational inference, we compared the regulatory network as predicted by our methods with published regulatory interactions observed in the model plant species, *Arabidopsis*. We combined three recently published *Arabidopsis* genome-wide gene regulatory networks (Sparks et al., 2016), including 2,914 regulatory interactions between 578 regulators and 717 targets. We mapped soybean genes to *Arabidopsis* genome and searched for predicted regulator-module interactions, which are also found in *Arabidopsis*. Among 54 interactions we predicted, five TF-module interactions and 13 TF-gene interactions from this dataset are also found in the *Arabidopsis* regulatory networks (Supplementary Table 7), providing support for the predicted gene regulatory networks. This small overlapping is expected because the current *Arabidopsis* interaction network (2914 interactions) is

only a tiny fraction of true interactions that happen *in-vivo* (see discussion).

Transcription factors regulate their target through binding of sequence specific motifs in the promoter regions of the target genes. To further validate the predicted regulatory networks, we performed promoter motif search. Among many programs that are available for motif discovery, the MEME suite contains the most comprehensive sets of programs that allows users to perform motif discovery, motif search and motif comparison. We used the MEME program to identify motifs in the promoter regions of genes in each gene module. We tested the enrichment of these newly discovered motifs and identified 101 motifs that are enriched in 36 modules, with each module having one or more enriched motifs. In our predicted interaction networks, there are 54 transcription factors that regulate 32 modules. We found that 21 out of these 32 modules contain one or more enriched motifs (Supplementary Table 8). These motifs are



putative binding sites of the 54 transcription factors that are regulating genes in these modules.

To further validate the results of the motif search, we compared our newly discovered motifs to a database of motifs generated by direct sequencing of binding sites of over 400 *Arabidopsis* TFs. The pattern of these motifs is represented by position specific weight matrices (PSWM) (Bailey et al., 2009). This comparison aims at testing whether any of the enriched motifs are similar to the binding motifs of the genes in the same gene family in *Arabidopsis*. For example (Table 2), we found that a bZIP transcription factor (Glyma.19G244800, whose most similar *Arabidopsis* gene is AT5G28770) is regulating module 57 in our predicted network. Our analysis found a GCCACGT motif enriched in the promoter regions of module 57 ( $p < 1.21e-3$ ). This motif is highly similar to the binding motif (GCCACGT) of an *Arabidopsis* bZIP transcription factor ( $p < 1.82e-10$ ). Four such examples are shown in Table 2. Among the 21 modules with enriched motifs in our predicted regulatory networks, we found

that 17 motifs are highly similar to the corresponding *Arabidopsis* motifs in the same TF gene family (Supplementary Table 8), providing strong support to the validity of our predictions.

## Regulatory Network Changes and Genes in Phytic Acid Metabolic Pathways

To understand the connections between transcription regulation and metabolic pathways that were altered in the *3mlpa* and *1mlpa* mutant lines, we downloaded the metabolic pathway annotation from the SoyCyc 7.0 database. We mapped genes in myo-inositol metabolism, stachyose metabolism and sucrose metabolism to different gene modules, because these metabolites are altered as a result of *lpa* mutations (Supplementary Table 3). We found that there were 64 genes involved in these metabolic pathways, which mapped to 34 gene modules (Supplementary Table 3). Twelve of the 64 genes were involved in stachyose (or raffinose family oligosaccharides (RFOs)) biosynthesis and eight were associated with inferred gene regulatory networks



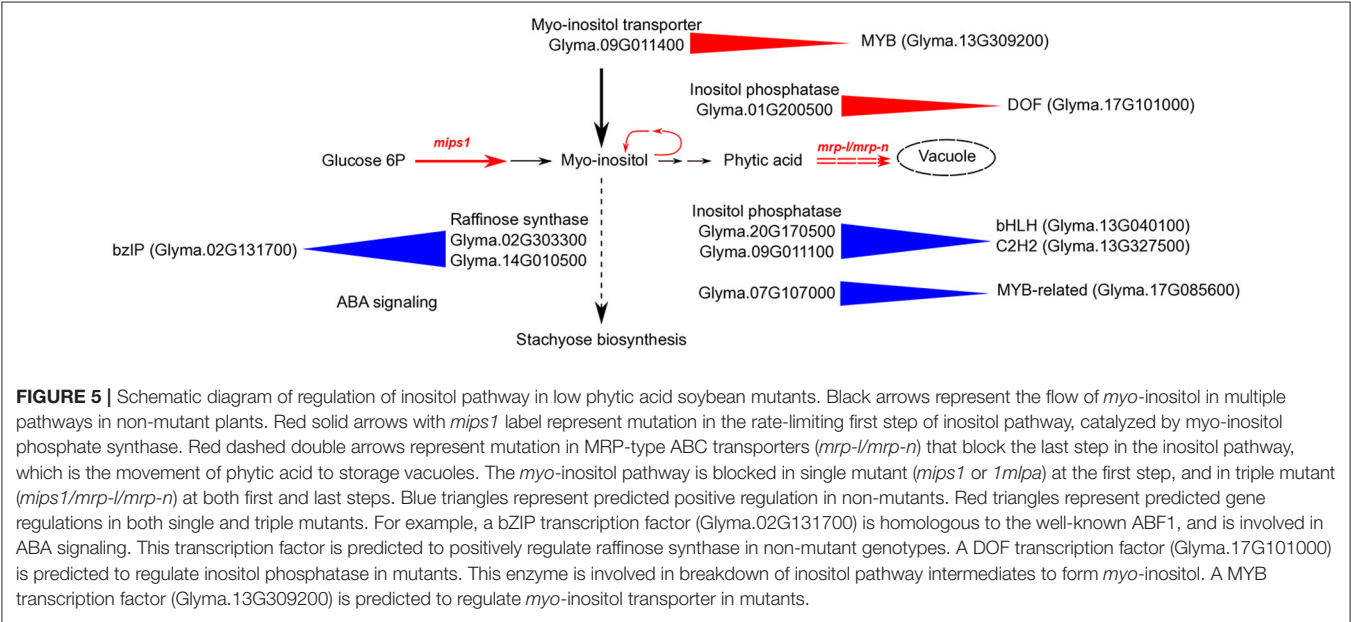
(Figure 5). In module 29, which is up-regulated in the first stage of seed development, two genes (Glyma.02G303300 and Glyma.14G010500, both encoding raffinose synthase / seed imbibition protein 1) were found to be regulated by a bZIP transcription factor (Glyma.02G131700). This bZIP transcription factor is homologous to *ABF1* (ABA response element-binding factor 1) in *Arabidopsis*, providing a potential connection between ABA response and stachyose biosynthesis (Figure 5). Some of the enzymes involved in *myo*-inositol biosynthesis are regulated similarly in mutant and non-mutant lines. For example, inositol-polyphosphate 5-phosphatase (Glyma.20G170500) and inositol-phosphate phosphatase (Glyma.09G011100) are found in module 16. Genes in this module are up-regulated during mid-stages of seed development. Module 16 is predicted to be regulated by two transcription

factors, a bHLH transcription factor (Glyma.13G040100) and a C2H2 transcription factor (Glyma.13G327500). The bHLH transcription factor is homologous to SPCH, which regulates stomatal lineage specification during embryo development (Danzer et al., 2015). Some regulatory interactions are only found in non-mutants. For example, genes from module 40 are predicted to be regulated by Glyma.17G085600, a MYB-related transcription factor, and this module contains two genes related to *myo*-inositol biosynthesis (Glyma.07G107000 encoding inositol-polyphosphate 5-phosphatase and Glyma.05G180600 encoding inositol-1-phosphate synthase). This module is highly expressed in the early stage of seed development and the MYB-related transcription factor is similar to *RSM1* in *Arabidopsis*, which has been found to be related to auxin signaling in early morphogenesis (Hamaguchi et al., 2008).

TABLE 2 | Motifs enrichment analysis.

TF name	TF family	Target module name	ATH TF name	Motif enrich adj p-value	MEME motif	DAP-seq motif	Motif similarity p-value
Glyma.19G244800	bZIP	57	AT5G28770	1.21E-03			1.82E-10
Glyma.03G042700	WRKY	55	AT2G38470	1.14E-03			4.14E-05
Glyma.05G110700	bHLH	35	AT4G37850	6.20E-03			8.09E-05
Glyma.17G101000	Dof	3	AT3G47500	3.79E-04			8.87E-05

"ATH TF name" is *Arabidopsis* transcription factor gene homologous to the soybean transcription factor gene. "Motif enrich adj p-value" is the enrichment p-value for each of the motif found in the promoter regions of predicted gene modules. "DAP-seq motif" is the most similar motif from DAP-seq database. "Motif similarity p-value" represents the significance of similarity between the motifs found in promoters in the gene modules and the DAP-seq motifs. All significantly enriched motifs and their corresponding DAP-seq motifs are provided in Supplementary Table 8.



Several target genes associated with regulation of phytic acid biosynthesis pathway matched with those discovered in co-expression network of developing maize kernel with low phytic acid content (Zhang et al., 2016). These predictions provide genotype-specific, testable hypotheses that may connect gene expression patterns with putative regulatory TFs and hormone regulations during different seed developmental stages.

## DISCUSSION

In this study, we performed computational inference of gene regulatory networks using data from developing soybean seeds from two mutants (*3mlpa* and *1mlpa*), with *lpa*-causing mutations, and the respective non-mutant siblings (3MWT and 1MWT). We identified co-expression gene modules with distinct and genotype specific expression patterns. These gene modules are also enriched with genes with various functional categories that are related to different stages of seed development. We identified transcription factors and their putative targets using supervised machine learning methods. Some of these transcription factors are differentially expressed only in non-mutants or only in mutants, suggesting that their regulations are lost or gained due to mutations. Many genes that encode enzymes in the metabolic processes of phytic acid, myo-inositol, sucrose, and stachyose and related oligosaccharides are found in these gene modules. Overall, our analysis provides a framework to connect transcription factors with genes in biological processes such as phytic acid metabolism, auxin-abscisic acid signaling and seed dormancy.

## Importance of Regulatory Network Inference

The predicted interactions provide a testable hypothesis for experimental validation using transgenic plants or ChIP-seq experiments. The results from this analysis can also be used to guide interpretation of other genomic mapping experiments such as genome wide association studies or quantitative trait loci analyses and to provide guidance for refining candidate gene lists. Soybean has a complex genome, which encodes over 4,500 putative transcription factors and over 46,000 coding genes. Although transcriptome data have become widely accessible for research in soybean and other crop species, it is still challenging to identify mechanistic connections between the observed transcriptome data with underlying regulatory networks. Differential gene expression analyses can be used to identify candidate genes that change under certain conditions or in specific mutants. However, in most situations, one still faces the problem of interpreting a long list of differentially expressed genes.

Our approach provides one alternative solution to this problem using well-developed machine learning methods to infer regulatory interactions. Our method implements the “community approach,” which has been shown to provide better performance than any individual method alone. The five computational methods are based on fundamentally different

statistical and mathematical formulations thus complement each other and provide a list of highly confident prediction results. Our approach successfully reduces the total number of candidate genes from over 10,000 genes that change in at least one comparison to 54 transcription factors, providing a much shorter list of key genes that can be focused on for validation experiments. However, combining five methods also limited the total number of predicted interactions (54 predicted interactions), because each method predicts some interactions that are not predicted by other methods.

We validated the predicted interactions using *Arabidopsis* gene regulatory networks. *Arabidopsis* is the model plant species, which provides most gene regulatory network information among all plant species. Although soybean and *Arabidopsis* diverged over 120 million years ago, key genes in metabolic pathways and signaling networks are conserved in both species (Jung et al., 2000; Hegeman et al., 2001; Le et al., 2010; Xu et al., 2011; Leite et al., 2014; Gerrard Wheeler et al., 2016), which is expected for a physiological process as conserved as seed development. Therefore, one can expect some regulatory interactions to be conserved between these two species. In fact, 13 interactions predicted by our method are also found in the *Arabidopsis* gene regulatory networks (Supplementary Table 7). The *Arabidopsis* genome contains approximately 2200 TFs and more than 27000 genes (Jin et al., 2015, 2017; Cheng et al., 2016). There are more than 59 million possible interactions between these TF and genes. Although the number of biologically active interactions is probably <1% of all possible interactions, the total number of true interactions is still far more than the 2914 interactions that were used in this comparison. Therefore, it is not surprising that we found a small overlap between the predicted interactions and those identified in *Arabidopsis*. As more interactions will be identified in both plant species, we would expect such overlap to increase. In our analysis, 17 motifs from 21 enriched modules (Supplementary Table 8) similar to the motifs identified in *Arabidopsis* orthologous genes, indicating that the interactions identified in this study are likely to be conserved between the two species. The actual number of conserved interactions is likely to be underestimated, because the regulatory network from *Arabidopsis* is far from complete. Nevertheless, our results provide an important first step toward characterizing gene regulatory networks in soybeans and other crop species.

One would expect more transcription factors being active during the seed development process. To observe a larger number of predicted interactions, we also provide results that are predicted by four out of five methods (Supplementary Table 6). This can be further extended to include predictions from fewer methods. Although aggregating multiple methods has been shown to outperform individual methods, some predictions by a specific method can represent interactions that cannot be detected by other approaches. If a specific target gene or specific function is of interest, one can also use our method to generate a ranked list of all predictions for the target of interest and use predicted regulators as candidate genes.

## Regulation of *myo*-Inositol Metabolism in *3mlpa* Mutant Line

*Myo*-inositol is an essential signaling molecule with multifunctional properties including gene regulation, chromatin modeling, mRNA transport, signal transduction, cell death, pathogen resistance, vesicle trafficking, plasma membrane, and cell wall formation (Martin, 1998; Stevenson et al., 2000; Chen and Xiong, 2010). It is synthesized in a two-step pathway (Figure 5), where glucose-6-phosphate is first converted to inositol monophosphate (IMP), a rate-limiting step catalyzed by MIPS1 enzyme, followed by dephosphorylation of IMP to form *myo*-inositol (Loewus and Murthy, 2000). Upon synthesis, *myo*-inositol is utilized in and recycled from multiple metabolic pathways such as biosynthesis of phytic acid and RFOs (such as stachyose and raffinose), inositol and phosphatidylinositol intermediates, auxin-inositol conjugates and glucuronic acid.

The *mips1* mutation in *Arabidopsis* is associated with reduction in cellular *myo*-inositol levels and defects in early embryogenesis (Meng et al., 2009; Chaouch and Noctor, 2010; Chen and Xiong, 2010; Donahue et al., 2010). The *mips1* mutation in soybean (as in parent V99-5089, *1mlpa* of this study) is associated with decreased levels of phytic acid and RFOs such as raffinose and stachyose, increased levels of sucrose and low emergence. The soybean *mips1* mutants also displayed normal RFO phenotype upon application of exogenous *myo*-inositol (Hitz et al., 2002). It is likely that, similar to *Arabidopsis*, the *mips1* mutation in soybean reduces *myo*-inositol levels, preventing RFO biosynthesis in parent V99-5089 (or in *1mlpa*), as *myo*-inositol is one of the starting substrates in this pathway. Since *myo*-inositol is not consumed by synthesis of RFO (it is a necessary intermediate, but recycled along the pathway), it is possible that the concentration of *myo*-inositol is reduced to such a low level such that synthesis of RFOs is greatly reduced. Sucrose is consumed in stachyose synthesis; it is therefore possible that inhibition of this pathway by the absence of *myo*-inositol is causing the increase in sucrose levels in V99-5089, due to accumulation of unused sucrose substrate.

The *mrp-l/mrp-n* mutations resulted in reduced seed phytic acid and low emergence, but did not alter the RFOs composition. In addition, the *myo*-inositol content in *mrp-l/mrp-n* mutant increases during the seed development phase prior to maturation (Israel et al., 2011). This suggests that in the presence of *mrp-l/mrp-n* mutations, the *mips1*-associated decrease in RFOs composition is restored, despite the reduced *myo*-inositol production due to the *mips1* mutation. This supports the prevalent hypothesis that the lack of transporters in the *mrp-l/mrp-n* mutant may trigger hydrolysis of cytoplasmic phytic acid to form inositol intermediates and *myo*-inositol, hence elevating the *myo*-inositol levels and preventing phytic acid synthesis, diverting the metabolic pathways to *myo*-inositol production (Figure 5). Another possibility is that lack of intracellular *myo*-inositol or phytic acid triggers feedback regulation that up-regulates transporters which can import inositol from the outside of the seed, hence elevating *myo*-inositol levels. This is in agreement with increase in the expression of inositol transporter genes during early stages of seed development in *3mlpa* mutant (Redekar et al., 2015).

In the present study, the inositol transporter gene (Glyma.09G011400) is also up-regulated in *1mlpa*, a single *mips1* mutant. This common up-regulation of inositol transporter in *mips1* and in *mips1/mrp-l/mrp-n* mutants suggests that these mutations are triggering the same signaling pathway (Figure 5). This inositol transporter gene belongs to module 5 and is predicted to be regulated by a MYB transcription factor (Glyma.13G309200), which represents a promising target for experimental validation (Figure 5). These target genes are in agreement with those identified in developing maize kernel with low phytic acid content (Zhang et al., 2016).

We identified over 30 genes associated with *myo*-inositol metabolism in this network analysis (Supplementary Table 3). Some of these genes belong to the modules that are only regulated in mutants. For example, an inositol-polyphosphate 5-phosphatase (Glyma.01G200500) is predicted to be regulated by a DOF transcription factor (Glyma.17G101000). This enzyme catalyzes an intermediate step in converting phytic acid to *myo*-inositol. The DOF transcription factor is differentially expressed when comparing *3mlpa* with 3MWT at stages 1, 3, and 5 and is not differentially expressed in any other comparisons. This observation is consistent with our hypothesis that in *3mlpa* mutant, phytic acid is recycled to produce *myo*-inositol, which participates in RFO synthetic pathways.

## Regulation of Auxin-ABA Signaling and Seed Dormancy-Related Genes in *lpa* Mutants

A main goal of this work was to elucidate the molecular basis of negative downstream impacts of *lpa* mutations. Here we found one possible component of this: that in soybean these mutations greatly impact phytohormone pathways. In addition to *myo*-inositol, the importance of phytohormones such as auxin and ABA, in seed development is well-documented (Locascio et al., 2014). Auxin is a key hormone through all phases of seed development including embryogenesis, organ differentiation, endosperm formation, and seed maturation, whereas ABA is involved in onset and maintaining seed dormancy, and is active in the seed maturation and desiccation phases (Locascio et al., 2014). Cross signaling between auxin and ABA are involved in regulating seed dormancy and hence germination (Liu et al., 2013). Hormone signal transduction is controlled by regulating their biosynthesis, accumulation and distribution in different sections and stages of developing seeds by factors such as *myo*-inositol. The *Arabidopsis mips1* mutants have demonstrated defects in cotyledon development and reduced expression of basipetal auxin efflux carriers such as *PIN1* and *PIN2* (Chen and Xiong, 2010; Luo et al., 2011). The factors regulating cellular *myo*-inositol levels might therefore also regulate cross signaling of auxin and ABA. However, it is still unclear to what extent auxin-ABA cross signaling is disturbed in *lpa*-mutants.

In this study, nearly the entire spectrum of the auxin signaling pathway was identified and was differentially regulated in *1mlpa* and *3mlpa* mutants as compared to the corresponding wild types. We identified 188 auxin-related genes differentially expressed in one or more comparisons tested in this study

(Supplementary Table 3). These included genes involved in auxin metabolism, transport, signal transduction, and transcription regulation. We also observed 36 genes from the ABA signaling pathway differentially expressed in one or more comparisons. The *1mlpa* vs. 1MWT comparison identified down regulation of two ABA catabolism genes in *1mlpa* (Glyma.01G153300 and Glyma.09G218600, which remove active ABA), whereas, in *3mlpa* vs. 3MWT comparison, these two genes are up-regulated.

In summary, we identified potential candidate genes that may play a role in regulating inositol metabolism, auxin-ABA signaling and seed maturation-dormancy in low phytic acid soybean during seed development. Although follow up experiments are required to validate these findings, the comprehensive regulatory network and the computational analysis pipeline of this study has set the necessary groundwork for future hypothesis driven investigations.

## AUTHOR CONTRIBUTIONS

NR and MASM designed the experiment. NR performed the sequencing experiments. NR and SL analyzed the data. SL

developed the machine-learning pipeline. NR, SL, GP, VR and MASM wrote the paper.

## FUNDING

This project was funded by Bio-design and Bioprocessing Research Center, John Lee Pratt Fellowship Program, Agricultural Experiment Station Hatch Program, and Open Access Subvention Fund—all at Virginia Tech.

## ACKNOWLEDGMENTS

We would like to thank Virginia Tech's Advanced Research Computing servers and Translational Plant Sciences' MAGYK server for their support.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2017.02029/full#supplementary-material>

## REFERENCES

- Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208. doi: 10.1093/nar/gkp335
- Bassel, G. W., Lan, H., Glaab, E., Gibbs, D. J., Gerjets, T., Krasnogor, N., et al. (2011). Genome-wide network model capturing seed germination reveals coordinated regulation of plant cellular phase transitions. *Proc. Natl. Acad. Sci. U.S.A.* 108, 9709–9714. doi: 10.1073/pnas.1100958108
- Bowen, D. E., Souza, E. J., Guttieri, M. J., Raboy, V., and Fu, J. (2007). A low phytic acid barley mutation alters seed gene expression. *Crop Sci.* 47, S-149. doi: 10.2135/cropsci2006.07.0456tpg
- Chaouch, S., and Noctor, G. (2010). Myo-inositol abolishes salicylic acid-dependent cell death and pathogen defence responses triggered by peroxisomal hydrogen peroxide. *New Phytol.* 188, 711–718. doi: 10.1111/j.1469-8137.2010.03453.x
- Chen, H., and Xiong, L. (2010). myo-Inositol-1-phosphate synthase is required for polar auxin transport and organ development. *J. Biol. Chem.* 285, 24238–24247. doi: 10.1074/jbc.M110.123661
- Cheng, C.-Y., Krishnakumar, V., Chan, A., Schobel, S., and Town, C. D. (2016). Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *bioRxiv*.
- Danzer, J., Mellott, E., Bui, A. Q., Le, B. H., Martin, P., Hashimoto, M., et al. (2015). Down-regulating the expression of 53 soybean transcription factor genes uncovers a role for speechless in initiating stomatal cell lineages during embryo development. *Plant Physiol.* 168, 1025–1035. doi: 10.1104/pp.15.00432
- Donahue, J. L., Alford, S. R., Torabinejad, J., Kerwin, R. E., Nourbakhsh, A., Ray, W. K., et al. (2010). The *Arabidopsis thaliana* Myo-inositol 1-phosphate synthase1 gene is required for Myo-inositol synthesis and suppression of cell death. *Plant Cell* 22, 888–903. doi: 10.1105/tpc.109.071779
- Faith, J. J., Hayete, B., Thaden, J. T., Mogno, I., Wierzbowski, J., Cottarel, G., et al. (2007). Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol.* 5:e8. doi: 10.1371/journal.pbio.0050008
- Gerrard Wheeler, M. C., Arias, C. L., Righini, S., Badia, M. B., Andreo, C. S., Drincovich, M. F., et al. (2016). Differential contribution of malic enzymes during soybean and castor seeds maturation. *PLoS ONE* 11:e0158040. doi: 10.1371/journal.pone.0158040
- Glover, N. (2011). *The Genetic Basis of Phytate, Oligosaccharide Content, and Emergence in Soybean*. Doctor of Philosophy, Virginia Tech.
- Gonzalez-Morales, S. I., Chavez-Montes, R. A., Hayano-Kanashiro, C., Alejo-Jacuinde, G., Rico-Cambron, T. Y., de Folter, S., et al. (2016). Regulatory network analysis reveals novel regulators of seed desiccation tolerance in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 113, E5232–E5241. doi: 10.1073/pnas.1610985113
- Hamaguchi, A., Yamashino, T., Koizumi, N., Kiba, T., Kojima, M., Sakakibara, H., et al. (2008). A small subfamily of Arabidopsis RADIALIS-LIKE SANT/MYB genes: a link to HOOKLESS1-mediated signal transduction during early morphogenesis. *Biosci. Biotechnol. Biochem.* 72, 2687–2696. doi: 10.1271/bbb.80348
- Haury, A. C., Mordelet, F., Vera-Licona, P., and Vert, J. P. (2012). TIGRESS: trustful inference of gene regulation using stability selection. *BMC Syst. Biol.* 6:145. doi: 10.1186/1752-0509-6-145
- Hegeman, C. E., Good, L. L., and Grabau, E. A. (2001). Expression of D-myo-inositol-3-phosphate synthase in soybean. Implications for phytic acid biosynthesis. *Plant Physiol.* 125, 1941–1948. doi: 10.1104/pp.125.4.1941
- Hitz, W. D., Carlson, T. J., Kerr, P. S., and Sebastian, S. A. (2002). Biochemical and molecular characterization of a mutation that confers a decreased raffinose and phytic acid phenotype on soybean seeds. *Plant Physiol.* 128, 650–660. doi: 10.1104/pp.010585
- Huynh-Thu, V. A., Irrthum, A., Wehenkel, L., and Geurts, P. (2010). Inferring regulatory networks from expression data using tree-based methods. *PLoS ONE* 5:12776. doi: 10.1371/journal.pone.0012776
- Israel, D. W., Taliencio, E., Kwanyuen, P., Burton, J. W., and Dean, L. (2011). Inositol metabolism in developing seed of low and normal phytic acid soybean lines. *Crop Sci.* 51, 282–289. doi: 10.2135/cropsci2010.03.0123
- Jervis, J., Kastl, C., Hildreth, S. B., Biyashev, R., Grabau, E. A., Saghai-Marroof, M. A., et al. (2015). Metabolite profiling of soybean seed extracts from near-isogenic low and normal phytate lines using orthogonal separation strategies. *J. Agric. Food Chem.* 63, 9879–9887. doi: 10.1021/acs.jafc.5b04002
- Jin, J., He, K., Tang, X., Li, Z., Lv, L., Zhao, Y., et al. (2015). An Arabidopsis transcriptional regulatory map reveals distinct functional and evolutionary features of novel transcription factors. *Mol. Biol. Evol.* 32, 1767–1773. doi: 10.1093/molbev/msv058
- Jin, J., Tian, F., Yang, D. C., Meng, Y. Q., Kong, L., Luo, J., et al. (2017). PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* 45, D1040–D1045. doi: 10.1093/nar/gkw982



- Jung, W., Yu, O., Lau, S. M., O'Keefe, D. P., Odell, J., Fader, G., et al. (2000). Identification and expression of isoflavone synthase, the key enzyme for biosynthesis of isoflavones in legumes. *Nat. Biotechnol.* 18, 208–212. doi: 10.1038/72671
- Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. doi: 10.1186/1471-2105-9-559
- Le, B. H., Cheng, C., Bui, A. Q., Wagmaister, J. A., Henry, K. F., Pelletier, J., et al. (2010). Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. *Proc. Natl. Acad. Sci. U.S.A.* 107, 8063–8070. doi: 10.1073/pnas.1003530107
- Le, B. H., Wagmaister, J. A., Kawashima, T., Bui, A. Q., Harada, J. J., and Goldberg, R. B. (2007). Using genomics to study legume seed development. *Plant Physiol.* 144, 562–574. doi: 10.1104/pp.107.100362
- Leite, J. P., Barbosa, E. G., Marin, S. R., Marinho, J. P., Carvalho, J. F., Pagliarini, R. F., et al. (2014). Overexpression of the activated form of the AtAREB1 gene (AtAREB1DeltaQT) improves soybean responses to water deficit. *Genet. Mol. Res.* 13, 6272–6286. doi: 10.4238/2014.August.15.10
- Li, S., Pandey, S., Gookin, T. E., Zhao, Z., Wilson, L., and Assmann, S. M. (2012). Gene-sharing networks reveal organizing principles of transcriptomes in Arabidopsis and other multicellular organisms. *Plant Cell* 24, 1362–1378. doi: 10.1105/tpc.111.094748
- Liu, W., Mazarei, M., Peng, Y., Fethe, M. H., Rudis, M. R., Lin, J., et al. (2014). Computational discovery of soybean promoter cis-regulatory elements for the construction of soybean cyst nematode-inducible synthetic promoters. *Plant Biotechnol. J.* 12, 1015–1026. doi: 10.1111/pbi.12206
- Liu, X., Zhang, H., Zhao, Y., Feng, Z., Li, Q., Yang, H. Q., et al. (2013). Auxin controls seed dormancy through stimulation of abscisic acid signaling by inducing ARF-mediated ABI3 activation in Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 110, 15485–15490. doi: 10.1073/pnas.1304651110
- Locascio, A., Roig-Villanova, I., Bernardi, J., and Varotto, S. (2014). Current perspectives on the hormonal control of seed development in Arabidopsis and maize: a focus on auxin. *Front. Plant Sci.* 5:412. doi: 10.3389/fpls.2014.00412
- Loewus, F. A., and Murthy, P. P. N. (2000). myo-Inositol metabolism in plants. *Plant Sci.* 150, 1–19. doi: 10.1016/S0168-9452(99)00150-8
- Luo, Y., Qin, G., Zhang, J., Liang, Y., Song, Y., Zhao, M., et al. (2011). D-myo-inositol-3-phosphate affects phosphatidylinositol-mediated endomembrane function in Arabidopsis and is essential for auxin-regulated embryogenesis. *Plant Cell* 23, 1352–1372. doi: 10.1105/tpc.111.083337
- Maetschke, S. R., Madhamsheetiwar, P. B., Davis, M. J., and Ragan, M. A. (2014). Supervised, semi-supervised and unsupervised inference of gene regulatory networks. *Brief. Bioinformatics* 15, 195–211. doi: 10.1093/bib/bbt034
- Marbach, D., Costello, J. C., Kuffner, R., Vega, N. M., Prill, R. J., Camacho, D. M., et al. (2012). Wisdom of crowds for robust gene network inference. *Nat. Methods* 9, 796–804. doi: 10.1038/nmeth.2016
- Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., et al. (2006). ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* 7(Suppl. 1):S7. doi: 10.1186/1471-2105-7-S1-S7
- Martin, T. F. (1998). Phosphoinositide lipids as signaling molecules: common themes for signal transduction, cytoskeletal regulation, and membrane trafficking. *Annu. Rev. Cell Dev. Biol.* 14, 231–264. doi: 10.1146/annurev.cellbio.14.1.231
- Meng, P. H., Raynaud, C., Tcherkez, G., Blanchet, S., Massoud, K., Domenichini, S., et al. (2009). Crosstalks between Myo-inositol metabolism, programmed cell death and basal immunity in arabidopsis. *PLoS ONE* 4:e7364. doi: 10.1371/journal.pone.0007364
- Nagy, R., Grob, H., Weder, B., Green, P., Klein, M., Frelet-Barrand, A., et al. (2009). The Arabidopsis ATP-binding cassette protein AtMRP5/AtABCC5 is a high affinity inositol hexakisphosphate transporter involved in guard cell signaling and phytate storage. *J. Biol. Chem.* 284, 33614–33622. doi: 10.1074/jbc.M109.030247
- Ni, Y., Aghamirzaie, D., Elmarakeby, H., Collakova, E., Li, S., Grene, R., et al. (2016). A machine learning approach to predict gene regulatory networks in seed development in arabidopsis. *Front. Plant Sci.* 7:1936. doi: 10.3389/fpls.2016.01936
- O'Malley, R. C., Huang, S. C., Song, L., Lewsey, M. G., Bartlett, A., Nery, J. R., et al. (2016). Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell* 166, 1280–1292. doi: 10.1016/j.cell.2016.04.038
- Raboy, V. (1997). "Accumulation and storage of phosphate and minerals," in *Cellular and Molecular Biology of Plant Seed Development*, eds B. A. Larkins, I. K. Vasil (Dordrecht: Kluwer Academic Publishers), 441–477
- Raboy, V. (2007). The ABCs of low-phytate crops. *Nat. Biotechnol.* 25, 874–875. doi: 10.1038/nbt0807-874
- Ramsey, S. A., Klemm, S. L., Zak, D. E., Kennedy, K. A., Thorsson, V., Li, B., et al. (2008). Uncovering a macrophage transcriptional program by integrating evidence from motif scanning and expression dynamics. *PLoS Comput. Biol.* 4:e1000021. doi: 10.1371/annotation/e14ad837-e5ff-4bd5-a5f2-f32e784d75a2
- Redekar, N., Biyashev, R., Jensen, R., Helm, R., Grabau, E., and Maroof, S. M. A. (2015). Genome-wide transcriptome analysis of developing seeds from low and normal phytic acid soybean lines. *BMC Genomics* 16, 1074. doi: 10.1186/s12864-015-2283-9
- Saghai Maroof, M. A., and Buss, G. R. (2008). *Low Phytic Acid, Low Stachyose, High Sucrose Soybean Lines*. Google Patents.
- Saghai Maroof, M. A., Glover, N. M., Biyashev, R. M., Buss, G. R., and Grabau, E. A. (2009). Genetic basis of the low-phytate trait in the soybean line CX1834. *Crop Sci.* 49, 69. doi: 10.2135/cropsci2008.06.0362
- Schafer, J., and Strimmer, K. (2005a). An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics* 21, 754–764. doi: 10.1093/bioinformatics/bti062
- Schafer, J., and Strimmer, K. (2005b). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Stat. Appl. Genet. Mol. Biol.* 4, Article32. doi: 10.2202/1544-6115.1175
- Segal, E., Shapira, M., Regev, A., Pe'er, D., Botstein, D., Koller, D., et al. (2003). Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat. Genet.* 34, 166–176. doi: 10.1038/ng1165
- Sherlock, G. (2000). Analysis of large-scale gene expression data. *Curr. Opin. Immunol.* 12, 201–205. doi: 10.1016/S0952-7915(99)00074-6
- Shi, J. R., Wang, H. Y., Schellin, K., Li, B. L., Faller, M., Stoop, J. M., et al. (2007). Embryo-specific silencing of a transporter reduces phytic acid content of maize and soybean seeds. *Nat. Biotechnol.* 25, 930–937. doi: 10.1038/nbt1322
- Shi, J., Wang, H., Hazebroek, J., Ertl, D. S., and Harp, T. (2005). The maize low-phytic acid 3 encodes a myo-inositol kinase that plays a role in phytic acid biosynthesis in developing seeds. *Plant J.* 42, 708–719. doi: 10.1111/j.1365-313X.2005.02412.x
- Shi, J., Wang, H., Wu, Y., Hazebroek, J., Meeley, R. B., and Ertl, D. S. (2003). The maize low-phytic acid mutant lpa2 is caused by mutation in an inositol phosphate kinase gene. *Plant Physiol.* 131, 507–515. doi: 10.1104/pp.014258
- Sparks, E., Drapek, C., Gaudinier, A., Li, S., Ansariola, M., Shen, N., et al. (2016). Establishment of expression in the shortroot-scarecrow transcriptional cascade through opposing activities of both activators and repressors. *Dev. Cell* 39, 585–596. doi: 10.1016/j.devcel.2016.09.031
- Stevenson, J. M., Perera, I. Y., Heilmann, I., Persson, S., and Boss, W. F. (2000). Inositol signaling and plant growth. *Trends Plant Sci.* 5, 252–258. doi: 10.1016/S1360-1385(00)01652-6
- Stevenson-Paulik, J., Bastidas, R. J., Chiou, S.-T., Frye, R. A., and York, J. D. (2005). Generation of phytate-free seeds in Arabidopsis through disruption of inositol polyphosphate kinases. *Proc. Natl. Acad. Sci. U.S.A.* 102, 12612–12617. doi: 10.1073/pnas.0504172102
- Taylor-Teeple, M., Lin, L., de Lucas, M., Turco, G., Toal, T. W., Gaudinier, A., et al. (2015). An Arabidopsis gene regulatory network for secondary cell wall synthesis. *Nature* 517, 571–575. doi: 10.1038/nature14099
- Weber, H., Borisjuk, L., and Wobus, U. (2005). Molecular physiology of legume seed development. *Annu. Rev. Plant Biol.* 56, 253–279. doi: 10.1146/annurev.arplant.56.032604.144201
- Wilcox, J. R., Premachandra, G. S., Young, K. A., and Raboy, V. (2000). Isolation of high seed inorganic P, low-phytate soybean mutants. *Crop Sci.* 40: 1601. doi: 10.2135/cropsci2000.4061601x
- Xu, F., Meng, T., Li, P., Yu, Y., Cui, Y., Wang, Y., et al. (2011). A soybean dual-specificity kinase, GmSARK, and its Arabidopsis homolog, AtSARK, regulate leaf senescence through synergistic actions of auxin and ethylene. *Plant Physiol.* 157, 2131–2153. doi: 10.1104/pp.111.182899

- Xu, X. H., Zhao, H. J., Liu, Q. L., Frank, T., Engel, K. H., An, G., et al. (2009). Mutations of the multi-drug resistance-associated protein ABC transporter gene 5 result in reduction of phytic acid in rice seeds. *Theor. Appl. Genet.* 119, 75–83. doi: 10.1007/s00122-009-1018-1
- Zhang, S., Yang, W., Zhao, Q., Zhou, X., Jiang, L., Ma, S., et al. (2016). Analysis of weighted co-regulatory networks in maize provides insights into new genes and regulatory mechanisms related to inositol phosphate metabolism. *BMC Genomics* 17:129. doi: 10.1186/s12864-016-2476-x
- Zhu, M., Dahmen, J. L., Stacey, G., and Cheng, J. (2013). Predicting gene regulatory networks of soybean nodulation from RNA-Seq transcriptome data. *BMC Bioinformatics* 14:278. doi: 10.1186/1471-2105-14-278

**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 © 2017 Redekar, Pilot, Raboy, Li and Saghai Maroof. 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) or licensor 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.



# An Integrated “Multi-Omics” Comparison of Embryo and Endosperm Tissue-Specific Features and Their Impact on Rice Seed Quality

## OPEN ACCESS

### Edited by:

Glória Catarina Pinto,  
University of Aveiro, Portugal

### Reviewed by:

Koh Aoki,  
Osaka Prefecture University, Japan  
Monica Escandon,  
University of Aveiro, Portugal

### \*Correspondence:

Loïc Rajjou  
loic.rajjou@agroparitech.fr

### † Present Address:

Marc Galland,  
Department of Plant Physiology,  
Swammerdam Institute for Life  
Sciences, University of Amsterdam,  
Amsterdam, Netherlands  
Erwann Arc,  
Institute of Botany, University of  
Innsbruck, Innsbruck, Austria  
Sandrine Balzergue,  
Institut de Recherche en Horticulture  
et Semences (Institut National de la  
Recherche Agronomique,  
AgroCampus Ouest, Université  
d'Angers), Beaucouzé, France  
Joseph Tran and Halima Morin,  
Institute of Plant Sciences  
Paris-Saclay, Institut National de la  
Recherche Agronomique, Centre  
National de la Recherche Scientifique,  
Université Paris-Sud, Université  
d'Evry, Université Paris-Diderot,  
Sorbonne Paris-Cité, Université  
Paris-Saclay, Orsay, France  
Benoît Valot,  
UMR 6249 Chrono-Environnement,  
Haut de Chazal, Ambroise Paré,  
Besançon, France

Marc Galland<sup>1†</sup>, Dongli He<sup>2</sup>, Imen Lounifi<sup>1</sup>, Erwann Arc<sup>1†</sup>, Gilles Clément<sup>1</sup>,  
Sandrine Balzergue<sup>3†</sup>, Stéphanie Huguet<sup>3</sup>, Gwendal Cueff<sup>1</sup>, Béatrice Godin<sup>1</sup>, Boris Collet<sup>1</sup>,  
Fabienne Granier<sup>1</sup>, Halima Morin<sup>1†</sup>, Joseph Tran<sup>1†</sup>, Benoît Valot<sup>4†</sup> and Loïc Rajjou<sup>1\*</sup>

<sup>1</sup> IJPB, Institut Jean-Pierre Bourgin (INRA, AgroParisTech, CNRS, Université Paris-Saclay), Saclay Plant Sciences (SPS), Versailles, France, <sup>2</sup> Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, China, <sup>3</sup> IPS2, Institute of Plant Sciences Paris-Saclay (INRA, CNRS, Université Paris-Sud, Université d'Evry, Université Paris-Diderot, Sorbonne Paris-Cité, Université Paris-Saclay), POPS-Transcriptomic Platform, Saclay Plant Sciences (SPS), Orsay, France, <sup>4</sup> GQE-Le Moulon, Génétique Quantitative et Evolution (INRA Université Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay), PAPPPO-Plateforme d'Analyse Protéomique de Paris Sud-Ouest, Saclay Plant Sciences (SPS), Gif-sur-Yvette, France

Although rice is a key crop species, few studies have addressed both rice seed physiological and nutritional quality, especially at the tissue level. In this study, an exhaustive “multi-omics” dataset on the mature rice seed was obtained by combining transcriptomics, label-free shotgun proteomics and metabolomics from embryo and endosperm, independently. These high-throughput analyses provide a new insight on the tissue-specificity related to rice seed quality. Foremost, we pinpointed that extensive post-transcriptional regulations occur at the end of rice seed development such that the embryo proteome becomes much more diversified than the endosperm proteome. Secondly, we observed that survival in the dry state in each seed compartment depends on contrasted metabolic and enzymatic apparatus in the embryo and the endosperm, respectively. Thirdly, it was remarkable to identify two different sets of starch biosynthesis enzymes as well as seed storage proteins (glutelins) in both embryo and endosperm consistently with the supernumerary embryo hypothesis origin of the endosperm. The presence of a putative new glutelin with a possible embryonic favored abundance is described here for the first time. Finally, we quantified the rate of mRNA translation into proteins. Consistently, the embryonic panel of protein translation initiation factors is much more diverse than that of the endosperm. This work emphasizes the value of tissue-specificity-centered “multi-omics” study in the seed to highlight new features even from well-characterized pathways. It paves the way for future studies of critical genetic determinants of rice seed physiological and nutritional quality.

**Keywords:** rice, seed, endosperm, embryo, multi-omics, translation, starch,utelins

## INTRODUCTION

Seeds of high nutritional and physiological quality are essential for the benefit of mankind. Seeds are also at the forefront for the preservation of biodiversity through the plant conservation strategies in seed banks (Wyse Jackson and Kennedy, 2009; Westengen et al., 2013; Hay and Whitehouse, 2017). The quality of seeds comprises physiological, ecological and nutritional traits for agriculture, agroecology, and agro-food system. In terms of botanical provenance, the mature seed of Angiosperms is a patchwork of maternal and filial tissues (Walbot and Evans, 2003; Olsen, 2004; Nowack et al., 2010; Lafon-Placette and Kohler, 2014). The double fertilization of the haploid egg cell and diploid central cell by the two haploid sperm cells gives rise to a diploid embryo and to a triploid endosperm respectively. In addition, the seed coat (testa) is composed of several cell layers coming from the mother plant ovule and ovary. In cereals (*Poaceae*), the embryo is composed of the embryonic axis surrounded by a single cotyledon (scutellum) and it will form the future seedling. The mature endosperm is composed of four differentiated regions: the endosperm transfer cell (ETC) region, the embryo-surrounding region (ESR), the aleurone layer (AL), and the starchy endosperm (SE) cells (Olsen, 2004). While the ETC and AL cells remain alive at the end of cereal seed development, most of the ESR and SE cells have undergone programmed cell death (PCD) with characteristic DNA laddering and organelle degradation.

From an evolutionary point of view, one hypothesis stipulates that the endosperm would have derived from a supernumerary embryo originating from double fertilization and that would have evolved into an embryo-supporting structure (Friedman, 1998). Subsequently, the endosperm has evolved multiple roles related to the embryo. First, the endosperm protects the embryo from atmospheric oxygen that eventually leads to the formation of hydroperoxides and cell death (De Giorgi et al., 2015). In addition, critical cross-talk between abscisic acid (ABA) and gibberellin (GA) regulating seed development, size, dormancy or storage breakdown during germination are also the results of endosperm—embryo interactions (Fincher, 1989; Penfield et al., 2006; Bethke et al., 2007; Folsom et al., 2014; Yan et al., 2014; Bassel, 2016). Still, so far, few reports have addressed seed quality issues in terms of functional tissue specialization.

During seed maturation, orthodox seeds acquire desiccation tolerance (DT) and storability (longevity) as defined by the ability to withstand extreme water loss (to values around 0.1 g water per gram of dry weight) and to survive in the dry state (Hoekstra et al., 2001; Alpert, 2005). Among the mechanisms

that promote DT, the formation of a glassy cytoplasm and the subsequent decrease in molecular mobility is positively correlated with seed longevity (Buitink and Leprince, 2008). During the late phase of seed maturation, the accumulation of late embryogenesis abundant (LEA) proteins, heat shock proteins (HSPs), antioxidants and non-reducing sugars all together contribute to glassy state establishment (Boudet et al., 2006; Farrant and Moore, 2011; Kaur et al., 2016; Sano et al., 2016). How different tissues cooperate to establish seed DT and storability is still unclear especially in cereals with a persistent endosperm. Among the few existing studies, a proteomic analysis on maize *viviparous5* (*vp5*) mutant showed that LEA and HSPs were affected in the ABA-deficient embryos of the *vp5* mutants contrarily to the endosperm (Wu et al., 2014). Furthermore, a consequence of lipid degradation during storage, detoxification of the lipid peroxidation by-product malondialdehyde (MDA) by the rice aldehyde dehydrogenase 7 (*OsALDH7*) proved to be essential for DT (Shin et al., 2009). The null rice mutants of *OsALDH7* showed increased MDA resulting in reduced seed viability (Shin et al., 2009). Yet, the exact site of MDA generation during dry storage and seed aging remains unknown. The impact of active lipid degradation on rice seed quality was further reinforced by transgenic analysis of two rice lipoxygenases *OsLOX2* and *OsLOX3* since silencing and overexpression of these two enzymes acts in opposite directions on seed germination and longevity (Huang et al., 2014; Xu et al., 2015). Interestingly, the suppression of *OsLOX3* expression in the rice endosperm improved resistance to seed aging (Xu et al., 2015). The endosperm outer cuticle layer permeability can also preserve the embryonic components from oxidation by atmospheric oxygen (De Giorgi et al., 2015). A recent paper showed that during rice seed aging, the endosperm and embryo were differentially affected by seed aging in particular regarding glycolytic enzymes that decreased in abundance in the endosperm while increasing in the embryo (Zhang et al., 2016). Last but not least, seed storage proteins can also buffer the oxidative stress caused by seed aging as shown for Arabidopsis cruciferins (Rajjou et al., 2008; Nguyen et al., 2015). Despite these relevant studies, the majority of the molecular determinants of both DT and seed longevity in different seed tissues are still to be established especially in the embryo of cereal crops.

The generation of energy, digestion of seed storage proteins (SSPs), carbohydrates and/or triacylglycerols (TAGs) proved to be of paramount importance to obtain highly vigorous seeds. Within a few hours of imbibition, the seed embryo is rapidly resuming respiration thanks to the presence of a functional electron transport chain in undifferentiated pro-mitochondria (Ehrenschaft and Brambl, 1990; Logan et al., 2001; Howell et al., 2006). Later on, the differentiation of these pro-mitochondria in fully functional mitochondria participates to full metabolic resumption through establishment of the tricarboxylic acid cycle (TCA) (Lawlor and Vince, 2014). In contrast, it is less clear how the AL cells produce the requested energy to synthesize the large amounts of  $\alpha$ -amylases for starch degradation (Fincher, 1989). During barley seed development, the inner SE is mildly to severely hypoxic while the AL is not (Rolletschek et al., 2011). Still, the precise investigation of oxygen requirement and

**Abbreviations:** AL, aleurone layer; ALDH, aldehyde dehydrogenase; AsA, ascorbate; DHA, dehydroascorbate; DT, desiccation tolerance; eIF, eukaryotic initiation factor; ESR, embryo surrounding region; ETC, endosperm transfer cells; HAI, hours after imbibition; HSP, heat shock protein; LEA, late embryogenesis abundant; LOX, lipoxygenase; MDA, malondialdehyde; MDHA, monodehydroascorbate; MSR, methionine sulfoxide oxidase; NBB, naphthol blue black; PA, periodic acid; PCD, programmed cell death; PIMT, protein-L-isoaspartate O-methyltransferase; PLD, phospholipase D; QTL, quantitative trait locus; RFO, raffinose family oligosaccharides; ROS, reactive oxygen species; SE, starchy endosperm; SSPs, seed storage proteins; TAGs, triacylglycerols; TCA, tricarboxylic acid cycle; TIC, total ion current; XIC, extracted ion current.



consumption during germination in the different cereal seed tissues remains to be established.

The metabolic system for mRNA translation and protein synthesis is the most energy-requiring process in most organisms. Through the use of the translational inhibitor cycloheximide, it has been shown that translation of stored mRNAs is necessary and sufficient for both Arabidopsis and rice germination (Rajjou et al., 2004; Sano et al., 2012, 2013). On the other hand, DNA transcription by RNA polymerase II is necessary for seed vigor and seedling growth (Rajjou et al., 2004; Sano et al., 2012). In cereal seeds, the importance of mRNA translation in the AL during germination is crucial. Indeed, a strong synthesis and accumulation of  $\alpha$ -amylases, in response to gibberellic acid, participates to starch mobilization that in turns fuel the germinating embryo with oligosaccharides (Fincher, 1989). Interestingly, the scutellum also provides  $\alpha$ -amylases contributing to the amylolytic activity of the starch endosperm during germination (Subbarao et al., 1998) suggesting a close relationship between endosperm tissues and embryo. Nevertheless, only few studies have addressed the precise content of the cereal endosperm in terms of translational machinery especially compared to that of the cereal embryo that was historically used as a model cell-free system to translate mRNAs *in vitro* (Takai and Endo, 2010).

As a model species for cereals with a well-annotated genome (International Rice Genome Sequencing Project, 2005), rice seeds are widely studied by taking advantage of “omics” approaches (Koller et al., 2002; Tarpley et al., 2005; Xu et al., 2008; Jiao et al., 2009; Wang et al., 2010; Lee and Koh, 2011; Nguyen et al., 2012; Xue et al., 2012). Most studies on cereal seed focused either on the isolated embryo (Howell et al., 2009; Kim et al., 2009; Han et al., 2014) or on the whole seed (Yang et al., 2007). In contrast, a small number of works were performed on both the embryo and endosperm in the same experiment (Gallardo et al., 2007; Sreenivasulu et al., 2008).

Here, we used an integrated “multi-omics” approach combining transcriptomics, label-free quantitative shotgun proteomics and gas chromatography coupled to mass spectrometry (GC-MS)-based metabolomics on dry mature rice seeds from a reference rice cultivar (*Oryza sativa ssp japonica* cv Nipponbare). The present work was firstly aimed at comparing the compartmentalization of nutritionally relevant pathways between the endosperm and embryo (nutritional quality). These pathways could be further improved through metabolic engineering based on knowledge of the fine composition of both embryo and endosperm. Secondly, we highlighted genes potentially important for seed storability and germination (agricultural quality). Thirdly, a targeted study of the factors associated with the seed storage compounds (starch and proteins) were analyzed carefully. Altogether, these exhaustive datasets emphasize determinants of rice seed quality in a tissue-specific manner.

## METHODS

### Rice Biological Material

Dry mature rice seeds (*Oryza sativa ssp japonica* cv Nipponbare) were harvested in September 2011 at the “Centre Français du Riz”

(Mas du Sonailler, Arles, France). At the lab, seeds were dehulled and dissected in one embryo (E0) and one endosperm (A0) fraction with a sharp scalpel. The white rice (SE) fraction was obtained thanks to a lab bench rice milling machine as previously described (Galland et al., 2014a). Dry weight was determined on 10 bulks comprising 10 rice seeds. The corresponding E0, A0, and SE were placed in a dry oven at 105°C for 48 h and weighted on a precision lab balance (XP204, Mettler-Toledo, France).

### Optical and Confocal Microscopy

Fixation, inclusion into historesin and cutting of 5  $\mu$ m semi-thin sections were done exactly as previously described (Galland et al., 2014a). Proteins and complex carbohydrates were revealed using a Periodic Acid (PA)—Schiff /Naphthol Blue Black (NBB) staining. Samples semi-thin sections (5  $\mu$ m) were first hydrolyzed 5 min in 1% periodic acid (w/v), rinsed with tap water and distilled water and then colored with Schiff’s reagent for 10 min in complete darkness. Subsequently, sample slices were washed with sulfurous water that contains 5% (w/v) sodium metabisulfite, 250 mM HCl and distilled water for 1 min before washing with tap water and distilled water. NBB staining was done using a pre-heated (65°C) working solution that contains 0.1% (w/v) NBB, 10% (v/v) acetic acid and distilled water in which samples were placed for 5 min. One final thorough washing was done in tap water and samples sections were finally placed in acetic acid (7%, v/v) for 1 min or more if PA-Schiff staining was too weak. Finally, samples sections were mounted between glass slides in glycerol (Histomount, National Diagnostics, UK) for visualization before imaging with Leica optical microscopy (Leica Zeiss Axioplan, Leica Microsystems, Wetzlar, Germany).

Neutral lipids were imaged by confocal scanning fluorescence microscopy (Leica TCS SP2, Leica Microsystems SAS, France) using the Nile red dye (Greenspan et al., 1985). For neutral lipid observation, 100  $\mu$ m wide sections of rice dry mature seeds were cut using a vibrating blade microtome (Leica VT1000 S, Leica Microsystems SAS, France) in sterile distilled water. Then, sections were quickly put on a glass slide in 100  $\mu$ l of a Nile Red solution that contained 0.1% Nile Red (w/v) in 50% glycerol. The cell walls were counterstained by adding 100  $\mu$ l of a Calcofluor solution that contained 1% Fluorescent Brightener 28 (w/v) in a carbonate/bicarbonate buffer pH 9.2. For Nile Red imaging, 488 nm was used for excitation and emission was collected between 593 and 654 nm. For Calcofluor imaging, 405 nm was used for excitation and emission was collected between 412 and 483 nm.

### Metabolome Analysis by Gas Chromatography Coupled to Mass Spectrometry (GC-MS)

For each tissue (i.e., embryo, endosperm), metabolite samples were obtained starting from three replicates of 100 rice seeds manually dissected in embryo and endosperm fractions. 100 embryos and 100 endosperms were grinded with mortar and pestle under liquid nitrogen for the embryos and with a Cyclotec™ 1093 Sample Mill (FOSS, Hillerød, Denmark) for the endosperms. All samples were lyophilized and around 20 mg dry weight (DW) of each sample were placed in 2 ml Safelock Eppendorf tubes (Eppendorf AG, Hamburg, Germany).

All analysis steps including extraction, derivatization, analysis, and data processing were adapted from the original protocol described by Fiehn et al. (2008) and following the procedure described by Avila-Ospina et al. (2017). The extraction solvent was prepared by mixing water:acetonitrile:isopropanol at the volume ratio 2:3:3 allowing to extract metabolites with a broad range of polarities. For derivatization step, N-methyl-N-trimethylsilyl-trifluoroacetamide (MSTFA; Sigma-Aldrich) was used in silylation procedure of metabolites. Samples were analyzed on an Agilent 7890A gas chromatograph coupled to an Agilent 5975C mass spectrometer. Raw Agilent datafiles were converted in NetCDF format and analyzed with AMDIS (Automated Mass Deconvolution and Identification System; <http://chemdata.nist.gov/mass-spc/amdis/>). A home retention indices/ mass spectra library built from the NIST, Golm, and Fiehn databases and standard compounds was used for metabolites identification. Peak areas were then determined using the QuanLynx software (Waters, Milford, USA) after conversion of the NetCDF file in MassLynx format.

## RNA Isolation and Microarray Analyses

Total mRNAs were isolated from three replicates of 100 embryos and 50 endosperms and hybridizations on the Affymetrix GeneChip® Rice Genome Array (Affymetrix, Santa Clara, CA, USA) were performed as previously described (Galland et al., 2014a). To obtain presence/absence calls for each probe, we normalized the CEL files by the MAS5 algorithm (Affymetrix). The CEL files were then normalized with the GC-RMA algorithm using the “gcrma” library available from the R Bioconductor suite of open-source softwares (Huber et al., 2015). To determine differentially expressed genes in the embryo and endosperm transcriptomes, we performed a usual two group *t*-test that assumes equal variance between groups. The raw *P*-values were adjusted by the Bonferroni method. We considered a gene as differentially expressed if adjusted-value is < 0.01. To establish the Pearson correlation, we plotted the embryo against the endosperm normalized mean probe intensity. All raw CEL files are available from the Gene Expression Omnibus under the accession GSE43780 (for the embryo: GSM1071216, GSM1071217, GSM1071204; for the endosperm: GSM1071199, GSM1071201, GSM1071210).

## Proteome Analysis

### Protein Extraction and In-Gel Digestion

For embryo protein extraction, three replicates of 50 embryos were ground in liquid nitrogen using mortar and pestle. Then, total soluble proteins were extracted at room temperature in 400 µl thiourea/urea lysis buffer (7 M urea, 2 M thiourea, 6 mM Tris-HCl, 4.2 mM Trizma® base (Sigma-Aldrich, Lyon, France), 4% (w/v) CHAPS) supplemented with 50 µl of the protease inhibitor cocktail Complete Mini (Roche Diagnostics France, Meylan, France). Then, 15 µl of dithiothreitol (DTT, 1 M, Sigma-Aldrich), 2 µl of DNase I (Roche Diagnostics) and 5 µl of RNase A (Sigma-Aldrich) were added to the sample. For endosperm protein extraction, three replicates of 5 endosperms were ground in liquid nitrogen using mortar and pestle. Then, total soluble proteins were extracted at room temperature in

1 ml thiourea/urea lysis buffer (same composition as above) supplemented with 35 µl of DTT, 2 µl DNaseI and 10 µl RNase A. Finally, protein extracts were let to agitate for 2 h at 4°C. All samples were then centrifuged at 20,000 g at 4°C for 15 min. The resulting supernatant was submitted to a second clarifying centrifugation as above. The final supernatant was kept and protein concentrations in the various extracts were measured according to Bradford (1976) using Bovine Serum Albumin as a standard.

Twenty-five microgram of embryo and endosperm soluble protein extracts (*n* = 3 biological replicates) were subjected to SDS-PAGE analysis with 10% acrylamide (**Figure S1**). Each lane was systematically cut in 16 bands and submitted to in-gel digestion with the Progest system (Genomic Solution) according to a standard trypsin protocol. Gel pieces were washed twice by successive separate baths of 10% acetic acid, 40% ethanol, and acetonitrile. They were then washed twice with successive baths of 25 mM NH<sub>4</sub>CO<sub>3</sub> and ACN. Digestion was subsequently performed for 6 h at 37°C with 125 ng of modified trypsin (Promega) dissolved in 20% methanol and 20 mM NH<sub>4</sub>CO<sub>3</sub>. The peptides were extracted successively with 2% trifluoroacetic acid (TFA) and 50% ACN and then with ACN. Peptide extracts were dried in a vacuum centrifuge and suspended in 20 µl of 0.05% TFA, 0.05% HCOOH, and 2% ACN.

### LC-MS/MS Analysis

Peptide separation by NanoLC was performed as described previously (Bonhomme et al., 2012). Eluted peptides were analyzed on-line with a Q-Exactive mass spectrometer (Thermo Electron) using a nano-electrospray interface. Peptide ions were analyzed using Xcalibur 2.1 with the following data-dependent acquisition parameters: a full MS scan covering 300–1,400 range of mass-to-charge ratio (*m/z*) with a resolution of 70,000 and a MS/MS step (normalized collision energy: 30%; resolution: 17,500). MS/MS Step was reiterated for the 8 major ions detected during full MS scan. Dynamic exclusion was set to 45 s. A database search was performed with X!Tandem (Craig and Beavis, 2004). Enzymatic cleavage was declared as a trypsin with two possible misscleavage. Cys carboxyamidomethylation was set to static modifications. Met oxydation was set as possible modifications. Precursor mass and fragment mass tolerance were 10 ppm and 0.02 Th, respectively. The 7th annotation of the Rice Genome Annotation Project database (Kawahara et al., 2013; 66,338 proteins) and a contaminant database (trypsin, keratins) were used. Only peptides with an *E*-value smaller than 0.1 were reported.

Peptide quantification was performed by extracted ion current (XIC) using MassChroQ software (Valot et al., 2011). A 5 ppm precision windows was set for XIC extraction. We eliminated the peptide ions not specific of a single protein and, since a peptide ion was detected several times in one biological sample, we summed the Total Ionic Current (TIC) area under peak corresponding to the same peptide ion. We also removed peptide ions that were not reliably detectable by keeping only peptide ions detected at least twice out of the three biological replicates. We obtained a final number of 34,179 and 11,824 peptide ions in the embryo and endosperm respectively corresponded to 2,099

and 786 non-redundant proteins. Since several peptide ions corresponded to the same protein, we summed the total peptide ions' TIC area to get the overall protein abundance and we then  $\log_2$ -transformed this protein abundance.

### Gene Singular Enrichment Analysis

Gene Singular Enrichment Analysis were performed using the Gene Ontology analysis toolkit provided by the AgriGO web resource (Du et al., 2010) with the Affymetrix (transcriptome) or the corresponding tissue proteome (2099 embryo or 786 endosperm proteins) as backgrounds. The  $p$ -values generated by a classical hypergeometric overrepresentation test were adjusted by the Yekutieli False Discovery Rate.

### Phylogenetic Analysis

The amino acid sequence of the 12 known glutelins and the putative new glutelin, namely Glu-X, were retrieved from the Rice Genome Annotation Project (Ouyang et al., 2007). Protein sequences were aligned with Clustal Omega (Sievers et al., 2011) with allowed gaps and a distance matrix computed (BLOSUM62 matrix). The corresponding phylogenetic tree was built using the Neighbor-joining method.

### Measurement of Translational Activity

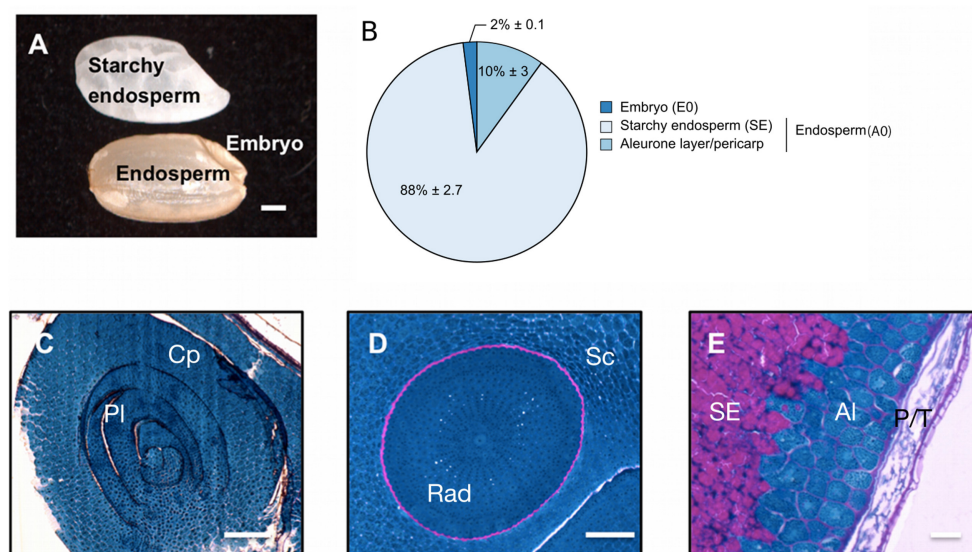
We imbibed three biological replicates of 20 isolated embryos and 5 embryoless endosperms in 4 ml of sterile distilled water with 50  $\mu\text{Ci}$  of [ $^{35}\text{S}$ ]-Met (PerkinElmer) at 30°C during 24 h in the dark. Samples were placed on filter papers to remove excess water and

grinded with mortar and pestle using liquid nitrogen. Proteins were then extracted according to previously published protocols (Rajjou et al., 2006). To avoid measuring the non-specific incorporation of radioisotopes into contaminants, we purified the total soluble proteins. In addition, dead seeds (autoclaved seeds) were used as a negative control in order to measure the background level related to non-specific incorporation of [ $^{35}\text{S}$ ]-Met. Finally, 10  $\mu\text{l}$  of protein extracts were added to 5 ml of scintillation liquid cocktail [Ecolite(+), MP Biomedicals, France]. Radioactivity was finally measured (3 biological and 3 technical replicates) using a liquid scintillation analyzer (Tri-Carb 2810TR, PerkinElmer, MA, USA) set between 5 and 100 keV with 10 min integration per sample.

## RESULTS

### Tissue Anatomy of the Rice Mature Seed

The pericarp, testa and AL were removed using a lab-polishing machine to obtain the inner part of the rice seed i.e., the starchy endosperm (Galland et al., 2014a; **Figure 1A**). On a dry weight basis, the embryo represents only 2% of the whole dry seed with the endosperm accounting for the remaining 98% (**Figure 1B**). Within the endosperm, the starchy endosperm represents 87.8% of the endosperm dry weight (**Figure 1B**). To describe the anatomy and content of the rice seed tissues, 5  $\mu\text{m}$  semi-thin sections of dry seeds were obtained and embedded in resin and stained with specific reagents (**Figures 1C–E**). The embryo appeared as a very cell-dense tissue rich in proteins with no



**FIGURE 1 |** Description of the dry mature rice seed. **(A)** The dry mature rice seed (*Oryza sativa* ssp. *japonica* cv Nipponbare) is composed of several tissues including the embryo, endosperm (pericarp, testa, aleurone layer and starchy endosperm) and inner starchy endosperm. Scale bar, 1 mm. **(B)** Dry weight per seed of the isolated embryo (E0), starchy endosperm and the aleurone layer/pericarp tissue. The average percentage of each seed tissue is indicated (average % per seed) along with its standard-deviation ( $n = 10$ ). The endosperm (A0) is the combination of the starchy endosperm (SE) and of the aleurone layer/pericarp tissue. **(C–E)** Proteins (blue) and complex carbohydrates (including starch, pink) were revealed using a Periodic Acid Schiff–Naphthol Blue Black staining on 5  $\mu\text{m}$  historesin-embedded semi-thick sections. Stained sections of the embryo shoot apical meristem **(C)**, longitudinal cut), embryo radicle **(D)**, transversal cut) and endosperm dorsal side **(E)** were visualized by optic microscopy. Scale bars represent 100  $\mu\text{m}$  in **(C,D)** and 25  $\mu\text{m}$  in **(E)**. Al, Aleurone layer; Cp, coleoptile; Pl, plumule; P/T, pericarp/testa; Rad, radicle; Sc, scutellum; SE, starchy endosperm.



complex sugars detectable except in cell walls and around the radicle (**Figures 1C,D**). In contrast, the endosperm appeared as heterogeneous tissue displaying a marked differentiation between the inner starchy endosperm (rich in starch and storage proteins) and the living AL (visible nuclei and numerous protein bodies) (**Figure 1E**). Lipids were detected by Nile Red tissue staining in both rice endosperm and embryo (**Figures 2A–C**), with local enrichments in the aleurone/subaleurone layers (**Figure 2B**) and scutellum epidermis (**Figure 2C**). These cytological observations show the important degree of compartmentation within the dry mature rice seed.

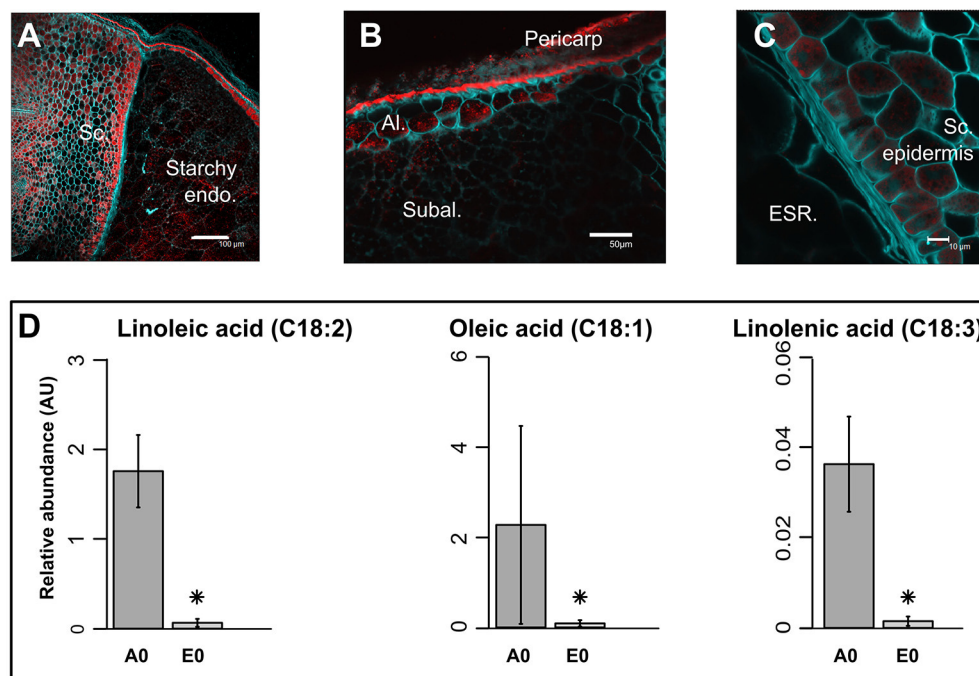
## Metabolic Composition of the Rice Mature Seed Compartments

Metabolomic data were generated from both embryo and endosperm (Table S1). Thus, 124 unique metabolites were identified at least once either in the rice seed embryo or endosperm and most of them (i.e., 117) were detected in all seed compartments (**Figure 3A**). Indeed, we only identified six embryo-specific metabolites i.e.,  $\gamma$ -tocopherol, feruloylquinic acid, maltotriose, adenosine-5-P and two galactinol isomer ( $m/z$  equal to 204 and 433, **Table 1**). In contrast, ascorbate was detected only in the endosperm (**Table 1**). A quantitative analysis was performed on the 117 common metabolites detected in both seed compartments. The abundances of the metabolites were normalized according to the dry weight of the embryo and

endosperm and a differential analysis performed that revealed 72 differentially accumulated metabolites ( $p < 0.05$ , **Table 1**). We found a strong correlation between the embryo and endosperm metabolite abundance per seed (**Figure 3B**). Thus, despite the differential ploidy and origin of the mature rice seed tissues, the composition in terms of primary metabolites is rather similar on a per seed basis. Nevertheless, most of these metabolites were more abundant in the endosperm such as unsaturated fatty acids: oleic acid (C18:1), linoleic acid (C18:2) and linolenic acid (C18:3) (**Figure 2D**). Only a few of them (e.g., raffinose, citrate,  $\alpha$ -tocopherol, glucaric acid, digalactosylglycerol) were significantly more abundant in the embryo (**Table 1**).

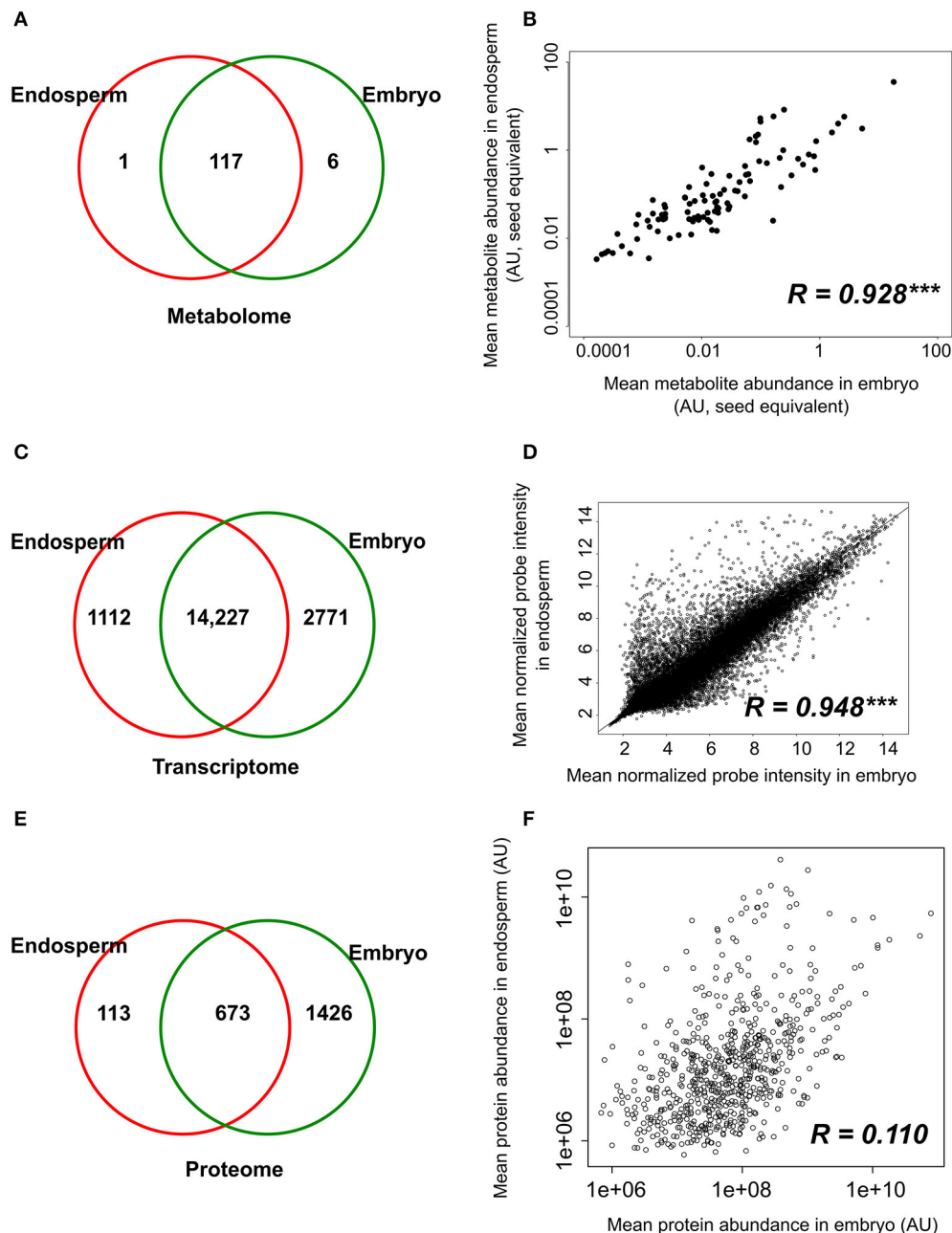
## Analysis of Long-Lived Stored mRNAs in the Embryo and Endosperm

The transcriptome of both the embryo and endosperm was analyzed using the Affymetrix Rice Genome Array and a dedicated workflow (Table S2, **Figure S2A**, Jung et al., 2008). Ambiguous and “absent” probe sets were removed which gave a final number of 15,339 and 16,998 detectable probe sets in the endosperm and embryo, respectively (equivalent to 12,964 and 14,150 unique genes representing 33 and 36% of the total genes) (Table S2C). The existence of a large overlap with 14,227 probe sets commonly detected in both endosperm and embryo was highlighted by these data (**Figure 3C**). Furthermore, a high significant correlation ( $R = 0.948$ ,  $p < 0.001$ ) was found



**FIGURE 2 |** Precise localization of oil bodies and compartmentation of fatty acids between rice seed tissues. **(A–C)** Lipids were visualized by confocal microscopy on fresh 100 µm sections of mature rice seeds. Fresh sections were stained by both Nile Red (red channel) to monitor neutral lipids and Calcofluor (green channel) to reveal cell walls. Scale bars are 100 µm in **(A)**, 50 µm in **(B)**, and 10 µm in **(C)**. Sc, scutellum; Endo, endosperm; Al, aleurone layer; Subal, subaleurone layer; ESR, embryo surrounding region. **(D)** Bar plots represent the mean seed equivalent lipid abundance ( $n = 3$ ). Asterisks indicate significant ( $p < 0.05$ , False Discovery Rate) different means. A0, endosperm; E0, embryo.





**FIGURE 3 |** Tissue-specificity and correlations between component tissue molecular apparatus. **(A,C,E)** Venn diagrams of the number of metabolites **(A)**, probe sets **(B)**, and proteins **(C)** detected in the dry embryo and endosperm. **(B,D,F)** Pearson correlations were calculated based on the metabolite level **(B)**, probe signal **(D)** and protein abundance **(F)** in the embryo and endosperm. Protein and metabolites are plotted on a  $\log_2$  and  $\log_{10}$  scale respectively. Transcript (probe sets) intensities are  $\log_2$  transformed during normalization and are plotted on non-transformed axis. Pearson correlation coefficients are indicated on each graph along with their significance level ( $^{***}p < 0.001$ ).

between the normalized probe intensities in the embryo and endosperm transcriptomes (**Figure 3D**). Yet, some tissue-specific transcripts were detected with 2771 embryo-specific (E0) probe sets (corresponding to 2613 single genes) and 1112 endosperm-specific (A0) probe sets (corresponding to 903 single genes, **Figure 3C**). Altogether, these results suggest that the endosperm

transcriptome is comparable qualitatively and quantitatively to that of the embryo.

The biological roles for the genes that were strongly (superior to median) and differentially regulated ( $p < 0.01$ ) between the endosperm and the embryo were then analyzed, which yielded to 787 and 1,921 probes (corresponding to 728 and 1746 single

**TABLE 1 |** Tissue-specific and differentially accumulated metabolites between embryo and endosperm of the dry mature rice seed.

Compound <sup>a</sup>	Class	Endosperm/embryo log <sub>2</sub> ratio <sup>b</sup>	FDR <sup>c</sup>
γ-tocopherol	Apolar	NA	Embryo
Maltotriose	Carbohydrate	NA	Embryo
Ascorbate	Organic acids	NA	Endosperm
Feruloylquinic acid	Organic acids	NA	Embryo
Adenosine-5-P	Purine_Pyrimidine	NA	Embryo
Galactinol_isomer_1	Sugar alcohol	NA	Embryo
Galactinol_isomer_2	Sugar alcohol	NA	Embryo
4-hydroxyproline	Amino acids	1.95	0.05
Alanine	Amino acids	0.88	0.04
beta-Alanine	Amino acids	1.76	0.03
3-cyano-alanine	Amino acids	3.31	0.03
Cysteine	Amino acids	1.77	0.05
4-Aminobutyrate (GABA)	Amino acids	2.29	0.01
Glutamine	Amino acids	3.00	0.01
Isoleucine	Amino acids	2.30	0.04
Leucine	Amino acids	2.21	0.03
Lysine	Amino acids	1.32	0.00
Ornithine	Amino acids	0.84	0.03
Phenylalanine	Amino acids	2.57	0.03
Pyroglutamate	Amino acids	0.97	0.03
Serine	Amino acids	2.05	0.03
Threonine	Amino acids	1.59	0.04
Tryptophan	Amino acids	1.20	0.01
Valine	Amino acids	2.39	0.04
α-Tocopherol	Apolar	−2.55	0.03
β-Sitosterol	Apolar	2.57	0.00
Campesterol	Apolar	1.72	0.00
Monopalmitin	Apolar	3.83	0.03
Monostearin	Apolar	4.62	0.02
Stigmasterol	Apolar	3.18	0.00
N-acetylmannosamine	Other	2.49	0.02
Arabinose	Carbohydrate	4.38	0.02
Cellobiose	Carbohydrate	1.95	0.03
Galactose	Carbohydrate	5.64	0.02
Glucopyranose	Carbohydrate	4.30	0.02
Mannose	Carbohydrate	1.54	0.01
Melibiose	Carbohydrate	1.49	0.01
Ribose	Carbohydrate	5.07	0.01
Sedoheptulose	Carbohydrate	3.56	0.02
Trehalose	Carbohydrate	1.53	0.00
Xylose	Carbohydrate	4.42	0.02
Raffinose	Carbohydrate	−0.76	0.02
Sucrose	Carbohydrate	0.98	0.01
Linoleic acid	Fatty acids	4.76	0.01
Linolenic acid	Fatty acids	4.61	0.02
Caffeate	Organic acids	4.73	0.03
Citrate	Organic acids	−1.24	0.01
Erythronate	Organic acids	3.57	0.05
Fumarate	Organic acids	3.20	0.05
4-hydroxybutanoate	Organic acids	5.34	0.01

(Continued)

**TABLE 1 |** Continued

Compound <sup>a</sup>	Class	Endosperm/embryo log <sub>2</sub> ratio <sup>b</sup>	FDR <sup>c</sup>
Glucaric acid	Organic acids	−2.69	0.01
Gluconate	Organic acids	1.96	0.03
Glycerate	Organic acids	3.80	0.03
Glycolate	Organic acids	4.57	0.03
Malate	Organic acids	2.77	0.01
Maleate	Organic acids	3.90	0.01
Malonate	Organic acids	2.35	0.05
Nicotinate	Organic acids	2.11	0.03
Phosphate	Organic acids	1.15	0.01
Pipecolate	Organic acids	0.68	0.01
Salicylate	Organic acids	4.03	0.05
Sinapinate	Organic acids	3.90	0.03
Sinapinate-trans	Organic acids	4.01	0.01
Succinate	Organic acids	3.28	0.03
Threonate	Organic acids	1.50	0.00
Spermidine	Polyamine	2.76	0.01
Guanine	Purine_Pyrimidine	4.31	0.03
Uracil	Purine_Pyrimidine	3.62	0.04
Erythritol	Sugar alcohol	3.59	0.00
Glycerol	Sugar alcohol	4.19	0.03
Glycerol-2-P	Sugar alcohol	1.56	0.01
Glycerol-3-P	Sugar alcohol	0.56	0.01
Inositol_isomer_3	Sugar alcohol	2.15	0.03
Inositol_isomer_4	Sugar alcohol	1.59	0.01
Inositol_isomer_5	Sugar alcohol	2.66	0.03
Inositol-1-P	Sugar alcohol	3.15	0.03
Mannitol	Sugar alcohol	3.98	0.03
Digalactosylglycerol	Sugar alcohol	−0.61	0.00
Galactosylglycerol	Sugar alcohol	4.38	0.03

<sup>a</sup>See additional file 1 for compound identification details (m/z and retention index)<sup>b</sup>The average seed metabolite abundance in each tissue (endosperm or embryo) was used to calculate the log<sub>2</sub> ratio. A positive log<sub>2</sub> ratio indicates an endosperm-favored metabolite accumulation (negative for embryo-favored log<sub>2</sub> ratio).<sup>c</sup>For each metabolite, a Student t-test followed by a False Discovery Rate (FDR) correction was computed. This table contains only metabolites with a  $p < 0.05$ . Tissue-specific metabolites are indicated (e.g., “Embryo”).

genes, respectively) that showed a preferential accumulation in the endosperm and embryo, respectively. Among the GO terms enriched in the 787 endosperm-favored probe sets, the “serine-type endopeptidase inhibitor activity” category, which contains known rice allergenic proteins (RAL2-5) associated with α-amylase or trypsin inhibitory functions, was found (AgriGO, Du et al., 2010,  $p < 0.05$ , **Figure S3**). Concerning the 1,921 probes with an embryo-favored expression, highly overrepresented GO terms related to “ribosome biogenesis,” “translation,” “rRNA binding,” “ribosomal large and small subunit,” and “structural constituent of ribosome” was detected (**Figure S4**). It seems therefore that a large proportion of the embryo long-lived mRNAs will serve as the basis for translation, a conserved and essential process for seed germination (Rajjou et al., 2012).

## Proteome Analysis of the Rice Dry Mature Seed

Regarding seed biology, post-transcriptional and translational regulations add a significant level of complexity as exemplified by studies on developing *Arabidopsis* and *Medicago* seeds where numerous examples of delays between mRNA accumulation and protein synthesis have been documented (Gallardo et al., 2007; Hajduch et al., 2010; Verdier et al., 2013). Thus, to complete the description of each rice seed compartment, a proteomic analysis was carried out. The total soluble proteins of embryo and endosperm tissues were extracted and these samples were subjected to a label-free quantitative shotgun proteomic analysis (Figure S1, Figure S2B, S5; Table S3). A total of 2212 single proteins were identified of which only 30.4% (673 proteins) were common to both compartments thereby revealing 1426 embryo-specific and 113 endosperm-specific proteins (Figure 3E). These results showed that the embryo proteome is much more diversified than the endosperm proteome. Furthermore, the abundance of the 673 common proteins is poorly correlated (Figure 3F) contrasting with what was observed at the transcriptome and metabolome levels (Figures 3B,D). For each of the 673 common proteins, the endosperm to embryo protein  $\log_2$  ratio was calculated and we found 76 and 267 proteins with an endosperm-favored or embryo-favored abundance respectively ( $\log_2$  ratios superior to the median i.e., 1.7 and  $-2.9$  for the endosperm and embryo). By combining these proteins showing a tissue-favored profile with the tissue-specific proteins, we obtained a list of 189 endosperm and 1,693 embryo proteins that we subsequently analyzed using the AgriGO tool (Du et al., 2010). In the endosperm, several expected enriched GO terms were retrieved such as “carbohydrate metabolic process” or “plastid” related to starch biosynthesis (Figure S6). In addition, others interesting enriched GO terms such as “response to endogenous stimulus,” “cell wall,” and “vacuole” were highlighted. These functions are probably related to the developed vesicle trafficking occurring during endosperm development and PCD. In the embryo, a strong GO enrichment for biological processes related to “translation,” “embryonic development,” “post-embryonic development” but also “response to stress” was found (Figure S7). At the subcellular level, several key organelles related terms were also overrepresented such as those related to the “mitochondrion,” “plastid,” or “ribosome” (Figure S7). Since the distribution of the  $\log_2$  ratios was similar to a normal distribution (Shapiro-test  $p = 0.95$ , Figure S5G), a z-score analysis was also performed. This revealed 64 proteins as being differentially accumulated between the embryo and the endosperm ( $p < 0.05$ ; Table S3D). Among these 64 differentially accumulated proteins, 47 are more abundant in the endosperm with several classical proteins classically found enriched in the endosperm such as SSPs (glutelins, prolamins, globulin) and starch biosynthesis enzymes (Table 2). Among the 10 glutelins commonly detected to both seed tissues, all of them were significantly ( $p < 0.05$ ) more abundant in the endosperm (7–13 fold, Figure 4A). An additional protein (Os08g03410, called Glu-X), annotated as a putative glutelin, is preferentially accumulated in the embryo (11 fold, Figure 4A). Glu-X display high sequence homology to other proteins belonging to the glutelin GluA/B/C/D families

(Figures 4B,C). Furthermore, we were able to find one eukaryotic translation factor eIF4A-1 or the rice homolog of the MOTHER of FT and TFL1 as two endosperm-favored proteins not directly related to classical endosperm proteins (Table 2).

## Identification of Post-transcriptional and Translational Regulations in the Mature Seed

The strong seed tissue differentiation observed at the proteome level could be well connected to the delay between transcript and protein accumulation observed at the end of seed development (Gallardo et al., 2007; Hajduch et al., 2010; Arc et al., 2011; Verdier et al., 2013). We quantitatively compared the 673 proteins fold-change between seed tissues with that of their corresponding mRNA level. These 673 proteins matched with 672 non-redundant genes that were used to retrieve their corresponding probe sets among the 14,227 common probe sets (Figure 3C). The probe sets corresponding to the same gene were removed leading to 504 unique probe set/protein pairs (Table S4). Then, the endosperm to embryo  $\log_2$  ratios of these 504 probe set/protein pairs were compared (Figure 5). From this analysis, it was obvious that the majority of the 504 protein  $\log_2$  ratios were poorly correlated to their corresponding mRNA  $\log_2$  ratio. On the one hand, proteins with an endosperm-favored  $\log_2$  ratio ( $>0$ ) also have similar endosperm-favored mRNA  $\log_2$  ratios. On the other hand, the proteins favorably accumulated in the embryo ( $\log_2$  ratio  $<0$ ) show poorly concordant mRNA profiles. These results once again highlight the seed as a tissue with major strong post-transcriptional and translational regulations probably related to the presence of long-lived mRNAs and to the seed metabolism being brought to a halt (Galland and Rajjou, 2015).

## Measurement of Translational Activities from Isolated Embryos and Embryoless Endosperms

During cereal seed germination, both embryo and endosperm tissues actively transcribe RNA and translate both stored mRNAs and *de novo* synthesized mRNAs (Fincher, 1989). The presently used rice seeds exhibited high vigor ( $T_{50} = 16$  HAI;  $G_{\max} = 24$  HAI). The isolated rice embryos were also capable of germination as evidenced by coleoptile emergence upon 24 h of imbibition (Figure 6A). Thus, the translational activities from isolated embryos and embryoless endosperms were assessed. As known for decades, the cereal embryo is an efficient system for *in vitro* translation of mRNAs (Takai and Endo, 2010) while the aleurone layer is actively synthesizing starch-degrading enzymes during germination (Fincher, 1989). Accordingly, considering a single seed, the endosperm had a slightly higher significant translational activity by comparison to the embryo (Figure 6B). By contrast, relatively to total protein content, translational activity is significantly higher in the embryo than in the endosperm (Figure 6C). This result indicates that protein synthesis is more active in the embryo. In support of this statement, the proteomic data revealed that the embryo is better equipped with proteins involved in the translation initiation machinery compared with the endosperm (Figure 6D).

**TABLE 2 |** Proteins differentially accumulated between endosperm and embryo.

Protein	Description	Endosperm to embryo $\log_2$ ratio <sup>a</sup>	p-value <sup>b</sup>
Os02g06410.1	Putative SNF-1 related protein kinase	-8.13	0.01
Os12g29400.1	GRAM domain containing protein, expressed	-8.02	0.02
Os05g39250.1	PEBP (phosphatidylethanolamine-binding) family protein	-7.67	0.02
Os05g49440.2	DUF1264 domain containing protein	-7.62	0.02
Os03g15960.1	17.9 kDa heat shock protein	-7.49	0.02
Os03g14180.1	26.7 kDa heat shock protein	-7.47	0.03
Os02g32860.1	Poly [ADP-ribose] polymerase 3	-7.46	0.03
Os03g19290.1	Chloroplastic outer envelope pore protein	-7.42	0.03
Os03g04410.1	Aconitase	-7.41	0.03
Os05g37330.1	60S acidic ribosomal protein	-7.38	0.03
Os02g51750.1	Annexin	-7.32	0.03
Os10g30150.1	Universal stress protein	-7.22	0.03
Os03g06360.1	Late embryogenesis abundant protein D-34	-7.18	0.03
Os01g50910.1	Late embryogenesis abundant protein LEA_4 domain	-6.87	0.04
Os05g44340.1	Chaperone protein ClpB1, heat shock protein 101	-6.86	0.04
Os08g41390.1	Putative 70 kDa peptidylprolyl isomerase	-6.74	0.04
Os10g17280.1	ATP synthase gamma chain	-6.71	0.04
Os07g42490.1	Sucrose synthase 3	8.80	0.00
Os07g11630.1	LTPL163 LTP family protein	7.94	0.00
Os12g14070.1	70 kDa heat shock-related protein	7.93	0.00
Os05g26350.1	13 kDa prolamin PROLM4	6.93	0.00
Os05g41970.1	19 kDa globulin	6.74	0.00
Os11g47520.1	Xylanase inhibitor protein 2	6.74	0.00
Os06g09450.1	Sucrose synthase 4	6.62	0.00
Os02g14600.1	Glutelin GluB-7	6.53	0.00
Os02g32660.1	Starch branching enzyme	6.53	0.00
Os07g11410.1	Seed allergenic protein RAG2	6.19	0.00
Os01g44220.1	ADP-glucose pyrophosphorylase	6.08	0.00
Os06g51084.1	Starch branching enzyme 1	6.07	0.00
Os07g10570.1	13 kDa prolamin PROLM25	5.98	0.00
Os02g15090.1	Glutelin GluD-1	5.84	0.00
Os06g04200.1	Granule-bound starch synthase 1 (Waxy)	5.81	0.00
Os05g33570.1	PPDK1 Pyruvate phosphate dikinase 1	5.79	0.00
Os07g11650.1	LTPL164 LTP family protein	5.46	0.01
Os07g11380.1	Seed allergenic protein RAG2	5.33	0.01
Os07g10580.1	13 kDa prolamin PROLM26	5.29	0.01
Os07g11360.1	Seed allergenic protein	5.27	0.01
Os06g22060.1	6-phosphofructokinase pyrophosphate dependent	5.11	0.01
Os02g25640.1	Glutelin GluC-1	5.03	0.01

(Continued)

**TABLE 2 |** Continued

Protein	Description	Endosperm to embryo $\log_2$ ratio <sup>a</sup>	p-value <sup>b</sup>
Os02g15169.1	Glutelin GluB-1b	4.88	0.01
Os07g11510.1	Seed allergenic protein	4.82	0.01
Os07g34520.1	Isocitrate lyase	4.79	0.01
Os02g16820.1	Glutelin GluB-5	4.77	0.01
Os07g11330.1	Seed allergenic protein	4.68	0.01
Os04g40660.1	MA3 domain containing protein	4.59	0.01
Os06g31070.1	13 kDa prolamin PROLM24	4.57	0.01
Os03g55090.1	Starch phosphorylase	4.54	0.01
Os03g16440.1	Translocon at outer-envelope membrane of chloroplast	4.51	0.01
Os03g31360.1	Glutelin GluA-3	4.34	0.02
Os04g33150.1	desiccation-related protein	4.14	0.02
Os10g35010.1	Translocon at inner-envelope membrane of chloroplast	4.00	0.02
Os02g15150.1	Glutelin GluB-2	3.79	0.03
Os10g26060.1	Glutelin GluA-2	3.69	0.03
Os06g48750.1	eIF4A-1 Eukaryotic translation initiation factor 4A-1	3.56	0.03
Os02g50350.1	Dihydropyrimidine dehydrogenase	3.55	0.03
Os01g55690.1	Glutelin GluA-1	3.51	0.03
Os11g40530.1	LTPL162 LTP family protein	3.44	0.03
Os07g11900.1	13 kDa prolamin PROLM19	3.30	0.04
Os01g47410.1	Aspartic proteinase oryzasin-1	3.24	0.04
Os06g46284.1	Probable alpha-glucosidase	3.15	0.04
Os06g30370.1	OsMFT1 homologous to Mother of FT and TFL1	3.08	0.05
Os07g11310.1	LTPL166 LTP family protein	3.06	0.05
Os05g02060.1	Chloroplastic outer envelope pore protein	3.03	0.05
Os02g32030.1	Elongation factor	3.00	0.05

<sup>a</sup>The average seed protein abundance in each tissue (endosperm or embryo) was used to calculate the  $\log_2$  ratio. A positive  $\log_2$  ratio indicates an endosperm-favored metabolite accumulation (negative for embryo-favored  $\log_2$  ratio).

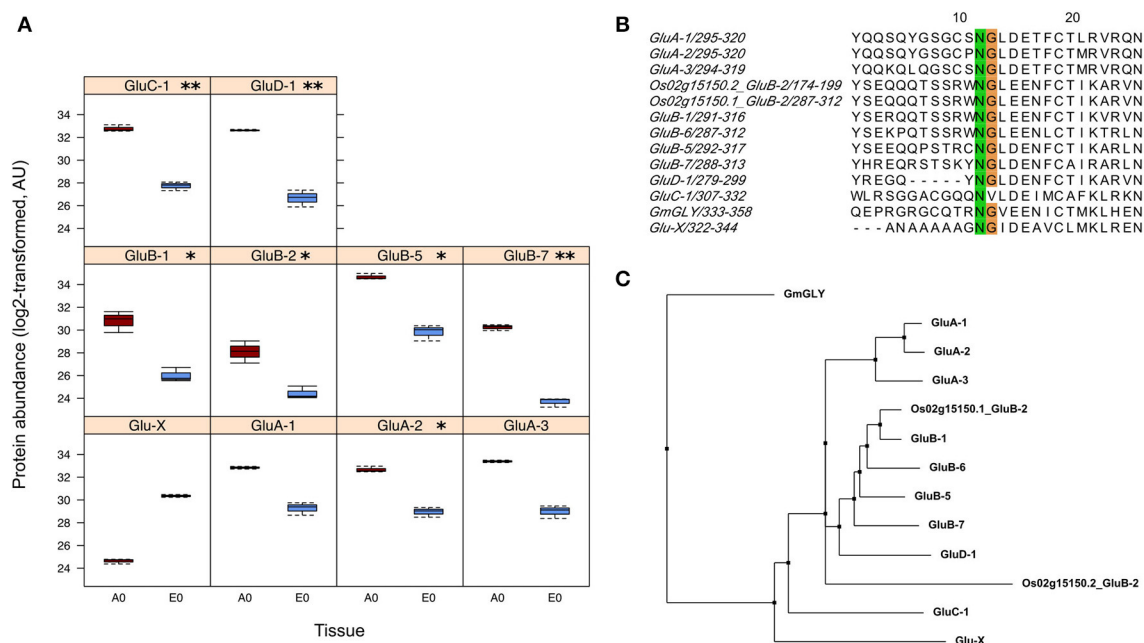
<sup>b</sup>P-value obtained from z-score analysis of endosperm to embryo  $\log_2$  ratios.

## DISCUSSION

### Origin and Roles of the Endosperm, a Key Tissue in Seed Biology

The supernumerary embryo origin of the endosperm is backed by several results in our study and that of others. As previously observed in the case of developing (Belmonte et al., 2013) or germinating Arabidopsis seeds (Penfield et al., 2006), the transcriptomes of the rice embryo and endosperm are highly similar both quantitatively and qualitatively (**Figures 3C,D**). In developing maize kernels (8 days after pollination), tissue-specific transcriptomics showed that the embryo transcriptome resembled that of aleurone cell layers but also that of other endosperm regions as Spearman correlations ranged from 0.73 (central starchy endosperm) to 0.80 (aleurone layer) (Zhan et al., 2015). In this latter study, the observed total number of maize endosperm-specific expressed genes (3140) is in the same order





**FIGURE 4 |** Identification of a putative new glutelin with a preferential embryo accumulation. **(A)** Boxplots of protein abundances (log<sub>2</sub> transformed) for the 10 glutelins that are commonly detected in the endosperm (A0, dark red) and the embryo (E0, blue). The Glu-X protein corresponds to a newly discovered glutelin which is the only glutelin more abundant in the embryo although not significantly. Significantly different means are indicated according to the z-score analysis with *p*-values inferior to 0.01 (\*\*) or 0.05 (\*). **(B)** Close-up view of the conserved cleavage site of glutelin precursor into acidic and basic subunits. The GluC-1 protein sequence does not exhibit the characteristic NG motif. **(C)** Phylogenetic tree built from the 12 glutelin protein sequences together with the newly putative Glutelin, namely Glu-X (neighbor joining, BLOSUM62 matrix)

of magnitude as ours (1112; **Figure 3C**). Similar conclusions can be drawn in for the embryo (2235 vs. 2771). Embryo and endosperm transcriptomes in developing and mature cereal caryopses could therefore be quite comparable to each other.

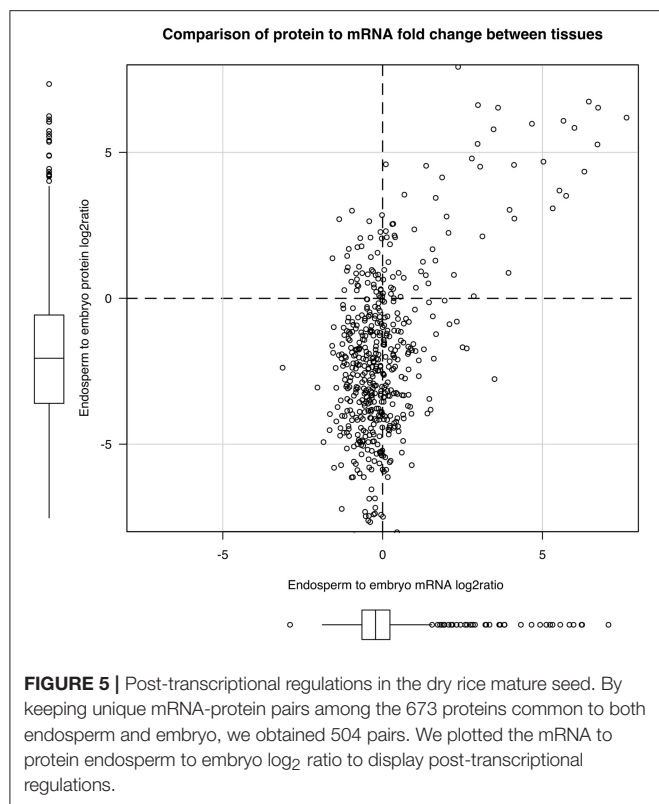
## Stored Reserves, Building Blocks, and Energy Source for Germination Starch

In rice, starch can account for nearly 70–85% of the total seed weight and is made in different proportions of linear amylose and ramified amylopectin (Bao et al., 2008). Historically, the degradation of starch during germination has been considered as a key mechanism that produced the oligosaccharides required for energy production during germination (Fincher, 1989). Here, the AMY3E  $\alpha$ -amylase protein (Os08g36900) was among the most abundant embryo-specific proteins (Table S3C) while a probe signal for the AMY3E mRNA was detectable in both embryo and endosperm (Table S2B). It is worth noting that no amylase activity could be measured in dry rice seeds (Guglielminetti et al., 1995) suggesting that the preformed embryo AMY3E enzyme is not functional at this stage. In the endosperm, only one  $\alpha$ -amylase protein was detected (Os01g51754) mildly ortholog of the *Vigna mungo* AMY1.1 enzyme (Table S3C). Finally, as reported before, no  $\beta$ -amylase protein was detected in rice seed compartments. Still, many  $\beta$ -amylase transcripts were found also suggesting that post-transcriptional and/or translational regulation is likely to occur for these enzymes.

Along with the presence of few starch-degrading enzymes, a number of starch biosynthesis enzymes were identified (**Figure 7A**). Starch biosynthesis originates from glucose-1-phosphate that is converted to ADP-glucose by ADP-glucose pyrophosphorylases (AGPases). ADP-glucose serves for  $\alpha$ -glucan chain elongation by starch synthases with participation of starch branching enzymes and starch debranching enzymes (Jeon et al., 2010). Most of the studies on rice starch biosynthesis have focused on the endosperm. Accordingly, a complete enzymatic machinery for starch biosynthesis and ADP-glucose transport (OsBT1-1) was detected in the endosperm (**Figure 7A**). While the presence of these enzymes is not surprising, the present work highlighted a complete starch synthesis enzymatic set already present in the quiescent embryo i.e., AGPS2a, SSIIB, GBSSII and Pho2 indicating that plastid is the main site of embryonic ADP-glucose synthesis (**Figure 7B**). In contrast, as a characteristic feature of graminaceous, ADP-glucose synthesis can occur in the endosperm thanks to a cytosolic ADP-glucose pyrophosphorylase (Beckles et al., 2001). Our results reveal a molecular specificity of each rice seed compartment for starch biosynthesis possibly related to starch synthesis in the embryo during germination (Han et al., 2014).

## Seed Storage Proteins (SSPs)

Glutelins are the major SSPs accumulated in rice corresponding to approximately 60–80% of the total proteins in the endosperm. During germination, these proteins are mobilized by proteases

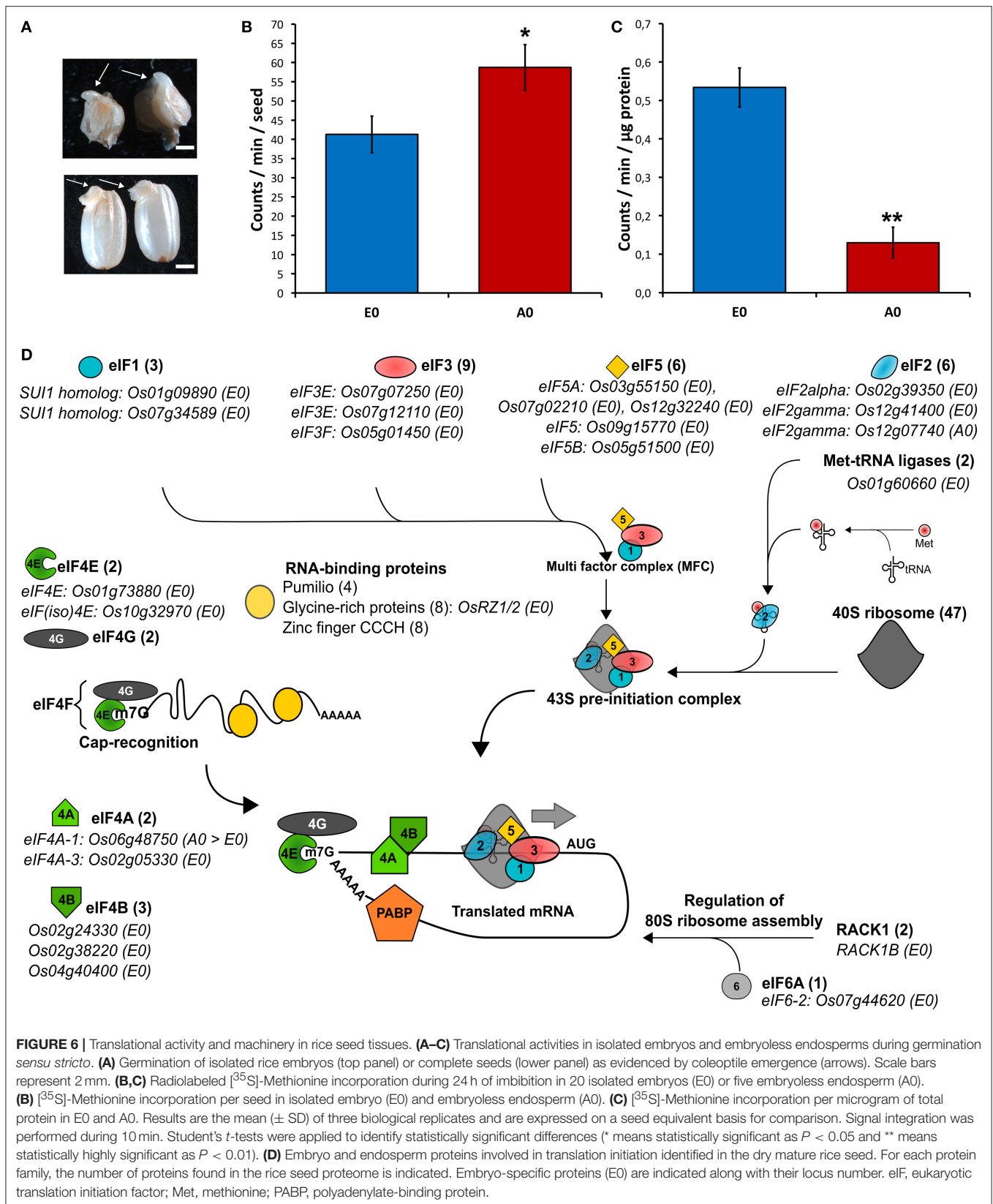


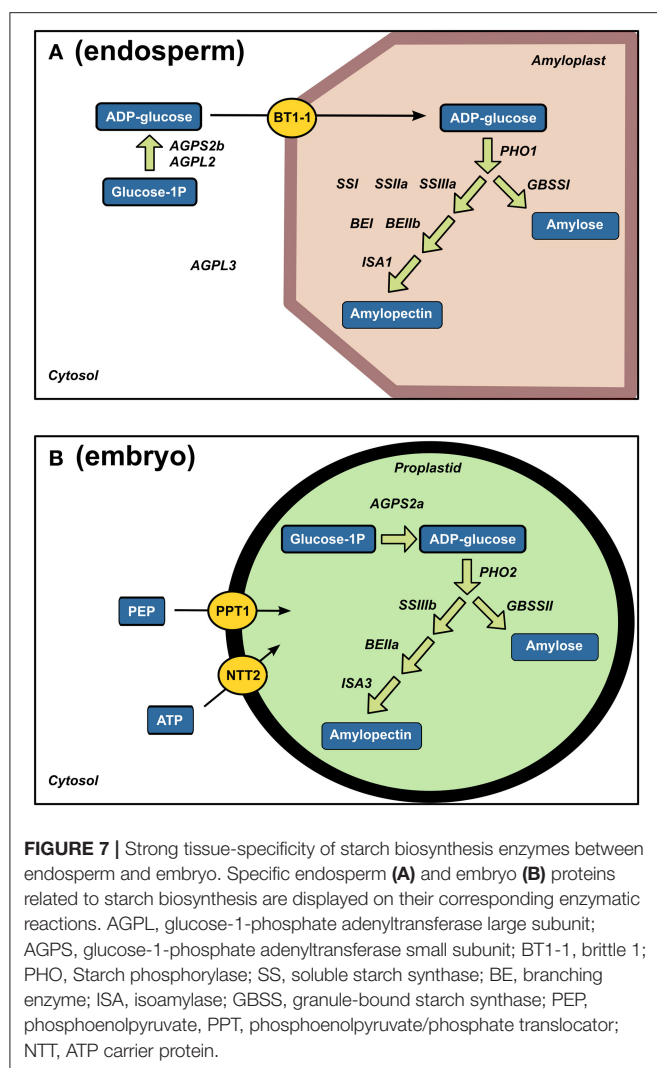
hereby releasing free amino acids that can readily be incorporated into new proteins requested for germination. Based on amino acid sequence similarity, previous studies established that 12 genes classified into GluA, GluB, GluC, and GluD families encode for rice glutelins (Kawakatsu et al., 2008). The present proteomic data evidenced 11 glutelin forms belonging to the GluA/B/C/D classification (Table S3C). Since the endosperm was separated from the embryo, it is remarkable to note that only two glutelin isoforms belonging to GluB class (i.e., Os02g15150.2 and Os02g15070.1) were specific of this storage tissue (Table S3C). In agreement with a storage function of the endosperm, among the 10 glutelins common to both seed tissues, all glutelins were significantly ( $p < 0.05$ ) more abundant in the endosperm (7–13 fold, **Figure 4A**). Unexpectedly, one protein (Os08g03410), annotated as a putative glutelin, is strongly accumulated in the embryo (11 fold, **Figure 4A**). This protein, hereby named Glu-X, is not classified in the glutelin GluA/B/C/D families but nevertheless presents the characteristic asparagine/glycine (NG) cleavage site specifically recognized by the vacuole-processing enzyme OsVPE (Wang et al., 2009) to process glutelin precursors into the corresponding acidic and basic subunits (**Figure 4B**; Kumamaru et al., 2010). This Glu-X protein is not closely related to the other glutelin families and could have a distinctive function in the embryo (**Figure 4C**). For instance, in Arabidopsis, cruciferins seed storage proteins protect the embryo from oxidative stress during seed aging (Rajjou et al., 2008; Nguyen et al., 2015).

## Lipids

In most cereal seeds, lipids generally account for only 2–3% of the dry weight (Barthole et al., 2012). In seeds, lipids are present in the form of TAGs that are stored in oil bodies corresponding to small vesicles composed of an inner TAG core surrounded by lipid monolayer containing dedicated proteins such as oleosins, caleosins and steroleosins (Murphy, 1993; Murphy et al., 2001; Jolivet et al., 2009). Lipids were detected in both rice endosperm and embryo with local enrichments in the aleurone/subaleurone layers (**Figure 2B**) and scutellum epidermis (**Figure 2C**). This precise localization is very similar to that of barley (Neuberger et al., 2008). Yet, the composition of unsaturated fatty acids such as oleic acid (C18:1), linoleic acid (C18:2) and linolenic acid (C18:3) were more abundant in the endosperm (**Figure 2D**).

The nutritional and organoleptic quality of the rice seed is highly dependent on polyunsaturated fatty acid peroxidation caused by lipoxygenases (LOXs). The Aldo-Keto Reductases (AKRs) protein family detoxifies a wide variety of lipid peroxidation compounds. Correspondingly, overexpression of Aldo-ketoreductase-1 from *Pseudomonas* strain (PsAKR1) in rice improved seed viability and germination vigor (Narayana et al., 2017). Two AKRs were specifically present in the embryo (Os04g26910 and Os05g38230) while one AKR was common to both embryo and endosperm (Os01g43090). Plant AKRs were recently proposed as potential breeding targets for developing stress tolerant varieties (Sengupta et al., 2015). In our proteomic data, we detected two rice lipoxygenases (OsLOX2, Os03g52860; OsLOX3, Os03g49350) specifically in the dry embryo in accordance with ancient biochemical results (Table S3C; Ida et al., 1983). Functional analysis showed that OsLOX2 and OsLOX3 negatively affect the germination performance of seeds submitted to artificial or natural aging (Huang et al., 2014; Xu et al., 2015). The higher occurrence of lipid hydroperoxides in aged seeds has been linked with a decrease in seed longevity (Sattler et al., 2004). In addition, the overexpression or silencing of OsLOX2 accelerates or slows germination *sensu stricto* (Huang et al., 2014). Rice LOX2 gene expression is induced upon germination presumably to degrade TAGs present in oil bodies and fuel seedling establishment. Therefore, TAGs that are stored in the AL oil bodies would release free fatty acids, fuel carbohydrate synthesis and energy metabolism in the AL and ETC cells during rice seed germination. It has been shown that TAG degradation occurs very early in the embryo and AL cells during the germination process (Clarke et al., 1983; Leonova et al., 2010). Following that hypothesis, the present work reveals in an unexpected way that several mRNAs encoding for glyoxylate cycle enzymes i.e., glyoxysomal malate dehydrogenase (MDH, Os12g43630 and Os05g50940), malate synthase (MLS, Os04g40990) and isocitrate lyase (ICL, Os07g34520) were more abundant in the rice endosperm (Table S2). Notably, ICL was also found as endosperm-favored at the protein level (**Table 2**). These glyoxylate cycle enzymes preferentially found in the rice endosperm might be associated with anoxia and stressful conditions (Lu et al., 2005). This would also explain the endosperm-favored accumulated of the pyruvate phosphate dikinase 1 (PPDK1, **Table 2**). In contrast, at





the protein level, the rice embryo appears favorably equipped with enzymes involved in glycolysis, tricarboxylic acid cycle and ATP synthesis (Figure 8). The degradation of membrane phospholipids during seed storage is also detrimental to seed quality (Devaiah et al., 2007). In particular, phospholipase D (PLD) enzymes that cleave membrane phospholipids to phosphatidic acid (PA) are proposed to be one the earliest event of deterioration. In Arabidopsis, silencing of the most abundant PLD enzyme, phospholipase D alpha 1 (AtPLDα1), improves seed longevity (Devaiah et al., 2007). Among the 17 *phospholipase D* (PLD) rice genes (Li et al., 2007), the *PLDα1* (OsPLDα1, Os01g07760) is the only one expressed and the corresponding enzyme is among the most abundant embryo-specific proteins (Table S3C). Among all rice PLD proteins, OsPLDα1 is closest relative of AtPLDα1 suggesting that their negative roles on seed longevity are probably conserved. In conclusion, the genetic manipulation of TAGs and phospholipid-related enzymes, in particular OsLOX2 and OsPLDα1, present in the rice embryo, have the potential to improve rice seed storability and organoleptic value.

## A Tissue-Specific Equipment Possibly Involved in Desiccation Tolerance and Seed Storability

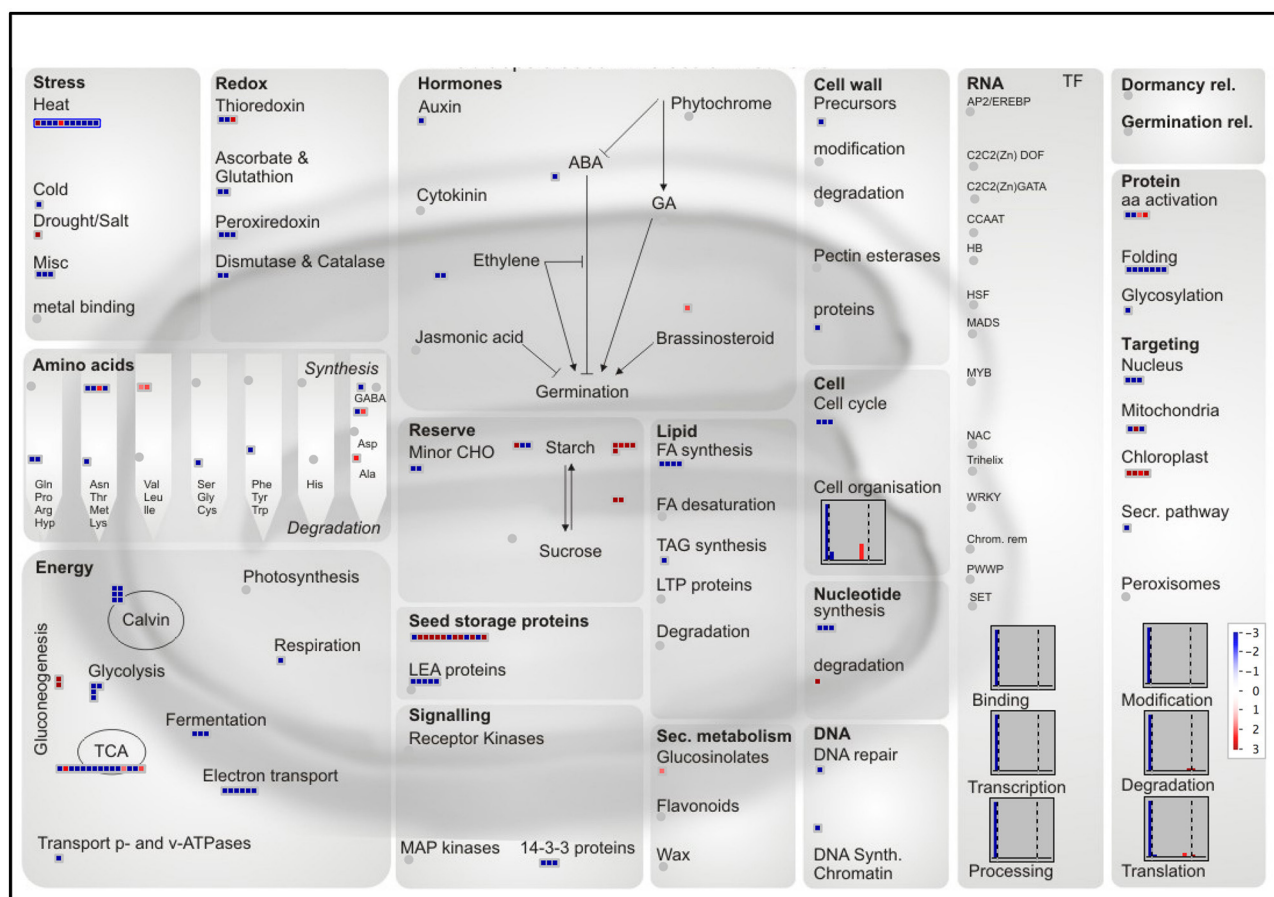
### Glassy State Establishment

Historically, rice seed longevity has been strongly linked to desiccation tolerance (Ellis and Hong, 1994). Among the mechanisms involved, the accumulation of non-reducing sugars (sucrose, trehalose) and raffinose family oligosaccharides (RFO) at the end of seed development converts the cellular cytoplasm into a “glassy state” that restricts molecule mobility and halt enzymatic reactions (Buitink and Leprince, 2004; Rajjou and Debeaujon, 2008; Farrant and Moore, 2011; Hand et al., 2011). RFO are galactosyl-sucrose carbohydrates that are formed by the sequential addition of galactose moieties by galactinol synthase, raffinose synthase, and stachyose synthase. In the present metabolomic data, sucrose, fructose, glucose, and raffinose were the major simple carbohydrates detected in the dry mature rice seed (Table 1 and Table S1). Furthermore, raffinose is preferentially accumulated in the embryo while non-RFO carbohydrates such as sucrose and trehalose are more highly accumulated in the endosperm (Table 1). Interestingly, a colocalization between QTL of longevity and QTL controlling oligosaccharide contents (sucrose, raffinose, and stachyose) has been pointed out in Arabidopsis and rice (Bentsink et al., 2000; Zhu et al., 2007). Together with raffinose, trehalose can also protect proteins and membranes damages induced by desiccation (Fernandez et al., 2010). It was recently found that the trehalose-6-phosphate phosphatase 7 (Os09g20390, *OsTPP7*) allele from an Indica cultivar was likely to be the underlying QTL for enhanced seed longevity in two Nipponbare near-isogenic lines (Sasaki et al., 2015). The rice genome harbors at least nine trehalose-6-phosphate synthase (*OsTPS*) and nine *OsTPP* genes (Fernandez et al., 2010). One natural candidate for favored trehalose synthesis in the endosperm aleurone layer could be *OsTPP10* (Os07g30160) whose transcript is only reliably detected in the endosperm (Table S2E) whereas *OsTPP8* protein (Os05g50940) was only detected in the embryo (Table S3C). Thus, the present work provides novel knowledge on the spatial regulation of genes involved in trehalose accumulation in rice seed and possibly related to desiccation tolerance and seed longevity. Thus, in the rice endosperm and aleurone layer in particular, the glassy state seems to be dependent on trehalose, sucrose and raffinose while in the embryo, it depends mostly on raffinose (Table 1).

### Protein Folding Protection by Molecular Chaperones

From the mapping of preferentially accumulated proteins, seed categories related to heat stress, protein folding and LEA proteins were quite noticeable (Figure 8). These categories contain protein chaperone roles such as the LEA proteins, (Wang et al., 2007; Hinch and Thalhammer, 2012), the small HSP (Sarkar et al., 2009; Waters, 2013), annexins (Clark et al., 2012), lipocalins (Grzyb et al., 2006), and ClpB chaperones. Most of these proteins were described to be involved in the maintenance of protein folding, prevent membrane aggregation and can also have a synergic effect with non-reducing sugars





**FIGURE 8 |** MapMan overview of differentially accumulated proteins with a preferential abundance in the endosperm or embryo. The  $\log_2$  ratios of the 335 proteins with a  $\log_2$  ratio superior to the median were mapped to a seed specific visualization (Joosen et al., 2011). A total of 241 proteins are visible. Red and blue colors represent endosperm and embryo-favored protein abundance respectively.

and RFO to promote glassy state establishment (Boudet et al., 2006; Rajjou and Debeaujon, 2008; Hand et al., 2011). Commonly related with desiccation and abiotic stress tolerance, LEA proteins are members of intrinsically disordered proteins in aqueous solution. They undergo desiccation-induced folding during cell drying suggesting that these proteins could carry out distinct functions under different water states. Rice comprises 34 LEA proteins encoding genes (Wang et al., 2007). Our proteomic analysis identified 12 LEA proteins detected in both embryo and endosperm and 9 only detected in the embryo (Table 3). LEA proteins were previously associated to seed longevity (Chatelain et al., 2012). It was remarkable to note that several LEA proteins, including the dehydrin family were detected exclusively in the embryo, and could be involved in dry storage survival (Table 3). Indeed, a previous study showed that dehydrin RAB18 was very abundant in Arabidopsis dry mature seeds. The abundance of this protein progressively disappeared in aged seeds (Rajjou et al., 2008). Furthermore, it has been demonstrated that down-regulation of seed-specific dehydrins reduced Arabidopsis seed survival in the dry state (Hundertmark et al., 2011). Out

of the 23 predicted sHSP proteins (Sarkar et al., 2009), the HSP17.4 was exclusively found in the endosperm (Table 3). This contrasts with the eight sHSP proteins exclusively found in the embryo of which the OsHSP18.2 (Os01g08860) is capable of protecting the Arabidopsis embryo during artificial aging (Kaur et al., 2015). Furthermore, all three common sHSP proteins (HSP16.9, HSP17.9 and HSP26.7) are preferentially more accumulated in the embryo (Table 3). Remarkably, the HSP16.9 protein was shown to stabilize rice soluble proteins from heat denaturation under *in vitro* conditions (Yeh et al., 1995). Altogether, these results support the finding that such proteins would primarily serve to protect the embryo against desiccation injuries during late maturation program. The present proteomic analysis also revealed several other categories of chaperone proteins such as annexins, lipocalins, and Clp (caseinolytic protease) chaperones which have never been characterized in cereal dry seeds. First, annexins are probably essential for seed longevity since the overexpression of a sacred lotus (*Nelumbo nucifera*) isoform in Arabidopsis proved to enhance seed viability under heat stress (Chu et al., 2012). In the present proteomic

**TABLE 3 |** Proteins involved in folding and chaperone functions found in the embryo and endosperm.

Protein	Family	Description	Specificity	Rank <sup>a</sup>	Endosperm to embryo log <sub>2</sub> ratio <sup>b</sup>
Os08g36150.1	ASA1	Heat shock protein 90 co-chaperone	Common	-	-3.9
Os05g44340.1	clpA/clpB	Chaperone protein ClpB1	Common	-	-6.9
Os10g42439.1	DnaJ	DnaJ homolog	Embryo	1,312	-
Os01g04370.1	HSP20	16.9 kDa class I heat shock protein 1	Common	-	-5.4
Os01g08860.1	HSP20	18.0 kDa class II heat shock protein	Embryo	74	-
Os02g52150.1	HSP20	24.1 kDa heat shock protein	Embryo	362	-
Os02g54140.1	HSP20	18.6 kDa class III heat shock protein	Embryo	49	-
Os03g14180.1	HSP20	Chloroplastic 26.7 kDa heat shock protein	Common	-	-7.5
Os03g15960.1	HSP20	17.9 kDa class I heat shock protein	Common	-	-7.5
Os03g16020.1	HSP20	17.4 kDa class I heat shock protein	Endosperm	75	-
Os03g16030.1	HSP20	18.1 kDa class I heat shock protein	Embryo	394	-
Os03g16040.1	HSP20	17.7 kDa class I heat shock protein	Embryo	195	-
Os04g36750.1	HSP20	23.2 kDa heat shock protein	Embryo	93	-
Os06g11610.1	HSP20	Mitochondrial 26.2 kDa heat shock protein	Embryo	1,210	-
Os11g13980.1	HSP20	21.9 kDa heat shock protein	Embryo	1,350	-
Os01g08560.1	HSP70	70 kDa Heat shock protein	Common	-	-4.5
Os01g62290.1	HSP70	70 kDa Heat shock protein	Common	-	-3.6
Os02g48110.1	HSP70	70 kDa Heat shock protein	Common	-	-2.5
Os02g53420.1	HSP70	70 kDa Heat shock protein	Common	-	-2.1
Os03g02260.1	HSP70	70 kDa Heat shock protein	Embryo	135	-
Os03g11910.1	HSP70	70 kDa Heat shock protein	Embryo	218	-
Os03g16860.1	HSP70	70 kDa Heat shock protein	Embryo	280	-
Os03g16920.1	HSP70	70 kDa Heat shock protein	Common	-	-3.5
Os03g60620.1	HSP70	70 kDa Heat shock protein	Common	-	0.6
Os05g08840.1	HSP70	70 kDa Heat shock protein	Common	-	-4.0
Os05g23740.1	HSP70	70 kDa Heat shock protein	Common	-	0.8
Os05g38530.1	HSP70	70 kDa Heat shock protein	Embryo	206	-
Os11g47760.1	HSP70	70 kDa Heat shock protein	Common	-	-2.3
Os12g14070.1	HSP70	70 kDa Heat shock protein	Common	-	7.9
Os04g01740.1	HSP90	90kDa heat shock protein	Embryo	92	-
Os06g50300.1	HSP90	Endoplasmic homolog	Common	-	-2.9
Os08g38086.3	HSP90	90 kDa heat shock protein	Endosperm	109	-
Os08g39140.1	HSP90	Heat shock protein 81-1	Embryo	129	-
Os09g29840.1	HSP90	90 kDa heat shock protein	Common	-	0.5
Os09g30412.1	HSP90	Heat shock protein 81-2	Embryo	551	-
Os09g30418.1	HSP90	Heat shock protein 81-3	Common	-	0.3
Os12g32986.1	HSP90	90 kDa heat shock protein	Embryo	179	-
Os01g06630.1	LEA	Late embryogenesis abundant LEA_5	Embryo	21	-
Os01g12580.1	LEA	Late embryogenesis abundant LEA_2	Common	-	-5.5
Os01g50700.1	LEA	Dehydrin Rab25	Common	-	-4.4
Os01g50910.1	LEA	Late embryogenesis abundant LEA_4	Common	-	-6.9
Os02g15250.1	LEA	Late embryogenesis abundant protein	Common	-	-4.5
Os02g44870.1	LEA	Dehydrin	Embryo	723	-
Os03g06360.1	LEA	Late embryogenesis abundant protein	Common	-	-7.2
Os03g20680.1	LEA	Late embryogenesis abundant protein 1	Common	-	-4.7
Os03g53620.1	LEA	Late embryogenesis abundant protein	Embryo	262	-
Os03g62620.2	LEA	Late embryogenesis abundant LEA_2	Common	-	-3.3
Os04g52110.1	LEA	Late embryogenesis abundant protein	Common	-	-5.0
Os05g28210.1	LEA	Embryonic abundant protein 1	Embryo	14	-
Os05g46480.1	LEA	Late embryogenesis abundant protein	Common	-	-5.9

(Continued)

TABLE 3 | Continued

Protein	Family	Description	Specificity	Rank <sup>a</sup>	Endosperm to embryo log <sub>2</sub> ratio <sup>b</sup>
Os05g50710.1	LEA	Late embryogenesis abundant LEA_2	Embryo	321	-
Os06g23350.1	LEA	Late embryogenesis abundant protein	Common	-	-4.5
Os08g23870.1	LEA	Late embryogenesis abundant LEA_1	Embryo	110	-
Os11g26570.1	LEA	Dehydrin	Common	-	-2.98
Os11g26750.1	LEA	Dehydrin Rab16D	Embryo	1,396	-
Os11g26780.1	LEA	Dehydrin Rab16B GN = RAB16B	Embryo	330	-
Os12g43140.1	LEA	Late embryogenesis abundant protein	Embryo	69	-
Os02g43020.1	TPR	Heat shock protein stress-induced TPLR_2	Common	-	-2.4
Os08g34150.1	Lipocalin	OsTIL-2 Temperature-induced lipocalin 2	Embryo	644	-
Os02g39930.1	Lipocalin	OsTIL-1 Temperature-induced lipocalin 1	Embryo	1,424	-
Os03g31300.1	clpA/clpB	Chloroplastic chaperone protein ClpB2	Common	-	-0.12
Os02g08490.1	clpA/clpB	Mitochondrial chaperone protein ClpB3	Embryo	165	-

<sup>a</sup>Rank of the protein respectively to the other tissue-specific proteins.

<sup>b</sup>The average seed protein abundance in each tissue (endosperm or embryo) was used to calculate the log<sub>2</sub> ratio. A positive log<sub>2</sub> ratio indicates an endosperm-favored metabolite accumulation (negative for embryo-favored log<sub>2</sub> ratio).

data, three annexins with one exclusively present in the embryo (Os09g23160) and one significantly more accumulated in the embryo (Os02g51750,  $p < 0.05$ ) were detected in the present study (Table 3). The last annexin (Os06g11800) was reported to be up-accumulated during germination suggesting a possible role on the embryo membrane dynamics (Yang et al., 2007). Secondly, lipocalins, a family of proteins that transport small hydrophobic molecules such as steroids, bilins, retinoids, and lipids, are classified in plants as temperature-induced lipocalins (TILs) and chloroplastic lipocalins (CHLs) (Charron et al., 2005). It has been demonstrated that both TILs and CHLs are involved in lipid protection, which is critical for stress adaptation. Two TIL proteins are predicted from the rice genome sequence (Charron et al., 2005) and they were detected in the mature rice embryo while the plastidial form OsCHL was undetectable in this tissue (Table 3). These results on the relative abundance of the OsTILs and OsCHL are consistent with those showed in Arabidopsis since the accumulation of AtCHL protein in the AtTIL KO mutant and *vice versa* suggests a functional overlap between these two lipocalin types (Boca et al., 2013). Interestingly, seed longevity is correlated with the accumulation of these proteins in Arabidopsis (Boca et al., 2013).

### Protein Repair Systems

Several enzymes involved in protein repair, were presently detected specifically in the rice embryo proteome. This was the case for three Methionine Sulfoxide Reductases (MSR) proteins namely MSRB5, MSRA2-1 and MSRA4 (Table S3C; Rouhier et al., 2006). MSRs are involved in the reversal of oxidized Met residues (Met sulfoxide, MetSO) in altered proteins thereby preventing aging-associated diseases in all organisms (Moskovitz, 2005). The MSRA4.1 is a plastidial enzyme potentially involved in oxidative stress resistance and that can repair free and protein-bound MetSO *in vitro* (Guo et al., 2009). MSR repair system in Medicago and Arabidopsis promote seed longevity (Châtelain et al., 2013). Secondly, the protein-L-isoaspartate O-methyltransferase can repair abnormal isoaspartyl occurring in damaged proteins (Thapar et al., 2001). In seeds, PIMT

are actively involved with the maintenance of seed viability in Arabidopsis (Ogé et al., 2008) and rice (Petla et al., 2016). In wheat, PIMT activity is very high in dry mature seeds, increase up to 4 h after imbibition and then decrease during subsequent germination (Mudgett and Clarke, 1994). Amongst the two rice PIMT genes, we found the OsPIMT2 (Os04g40540; Petla et al., 2016) among the most abundant embryo-specific proteins (rank #188). More precisely, this could be ΔOsPIMT2, a truncated yet functional version of OsPIMT2 (Petla et al., 2016). This rice PIMT protein is accumulated during the very late stages of seed development in relation due to the formation of aspartyl residues during desiccation (Petla et al., 2016).

### ROS Homeostasis

The control of Reactive Oxygen Species (ROS, e.g., H<sub>2</sub>O<sub>2</sub>) homeostasis during both desiccation and early germination is of paramount importance for seed vigor and longevity (Sattler et al., 2004; Bailly et al., 2008). Proteomic and Metabolomic results emphasized several mechanisms that could help the embryo to cope with desiccation-induced oxidative stress.

First, tocopherols and tocotrienols participate to seed longevity by limiting lipid peroxidation (Sattler et al., 2004). In our data, α- and γ-tocopherols were found to be preferentially, if not exclusively, accumulated in the embryo (Table 1). Several proteins involved in tocopherol biosynthesis pathway were specifically found in the embryo such as the 4-hydroxyphenylpyruvate dioxygenase (HPPD, Os02g07160; Table S3C), which is involved in the production of both plastoquinone and tocopherol essential for plant survival (Sano et al., 2016). In the same way, OsVTE1 protein (Os02g17650), that is responsible for γ-tocopherol synthesis, was specifically detected in the rice embryo (Table S3C). Finally, part of the same pathway, we found expression of Arabidopsis VTE2 (Os06g44840) and VTE3 (Os12g42090) homologs with a favored embryo gene expression (5 and 2.5 fold respectively, Table S2F).

Along with vitamin E, ascorbate is also a very important antioxidant molecule. Ascorbate (AsA) and dehydroascorbate (DHA) were specifically or favorably detected in the endosperm

(Table 1 and Table S1). Ascorbate and DHA can be degraded to threonate upon non-enzymatic reaction with  $H_2O_2$  or enzymatically. Interestingly, threonate is also present in high amounts in the endosperm compared to the embryo (Table 1). This suggests that a complete AsA to threonate pathway exist in the endosperm. In developing wheat kernels, ascorbate level decrease from mid- to final seed developmental stage and the ascorbate pool becomes progressively oxidized (Paradiso et al., 2012). From our data, it seems that ascorbate *de novo* synthesis could be restricted to the embryo since a putative mannose-1-phosphate guanylttransferase (Os03g11050) and the two GDP-mannose 3,5-epimerase 1 and 2 (GME1, GME2) are specifically detected in the embryo at similar abundances (Table S3C). In contrast, the ascorbate salvage pathway from monodehydroascorbate (MDHA) by MDHA reductase (MDHAR) is present in both rice endosperm and embryo (Table S3C). In addition, the conversion of DHA back to AsA is also possible thanks to DHA reductases (OsDHAR1, Os05g02530) present in both rice embryo and endosperm proteomes at similar levels (Table S3C). Ascorbate could interact with ABA metabolism and/or signaling to modulate seed germination ability. Indeed, exogenous application of low concentrations of ascorbate is able to rescue rice seed germination from abscisic acid treatment (Ye et al., 2012).

The present proteome reveals a wide diversity of antioxidant enzymes that are already present in the dry seed with an embryo-favored accumulation (Figure 8). These enzymes include 4 superoxide dismutases (SODCC1, SODCC2, SODA, SODCP, Table S3C), two embryo-specific catalases (CATA and CATB, Table S3C) and 11 embryo-specific peroxidases including several ascorbate peroxidase i.e., the cytosolic OsAPX2, the peroxisomal OsAPX4 and the stromal OsAPX7 (Table S3C). These results seem to argue in favor of a more abundant ROS detoxification enzymatic apparatus in the rice embryo.

## The Mature Seed Is the Crossroad of Post-transcriptional and Translational Regulations Essential for Germination Success

Studies in various species demonstrated that the developmental transition from a maturing to a germinating seed is the place of strong post-transcriptional and translational regulations (Gallardo et al., 2007; Hajdouch et al., 2010; Verdier et al., 2013; Galland et al., 2014b; Layat et al., 2014). In this study, we investigated that post-transcriptional and translational regulations occurring in both tissues of the rice seed at the end of its development (Figure 5). Since germination *sensu stricto* in both monocot and dicot seeds is only dependent on mRNA translation (Rajjou et al., 2004; Sano et al., 2012), we took a closer look at the translational machinery at the tissue-level.

### Stored mRNAs and the Translational Machinery

We were wondering whether these comparable translational activities in the endosperm and embryo relied on different translational machinery sets. Indeed, we could show that, in Arabidopsis germinating seeds, stored mRNAs were differentially

translated (Galland et al., 2014b) making selective mRNA translation a way to distinguish stored and neosynthesized mRNAs. For these reasons, proteins involved in translation and present in the embryo and endosperm were screened. Thus, 292 proteins related to mRNA translation processes (BIN 2.2.1–2.2.4) were identified in the embryo (Table S3F, Figure 6D). Specifically, 109 different ribosomal proteins could be identified. Initiation of translation is controlled by specific cap-dependent initiation factors (Roy and von Arnim, 2013). First, the 43S pre-initiation complex is formed through association between the 40S ribosomal subunit, a charged methionyl-tRNA and the eIF1, eIF2, eIF3 and eIF5 translation initiation factors. A complete set of the 43S pre-initiation complex was retrieved in the proteomic data (Table S3F). Out of the eight eIF3 protein subunits (B-C-D-E-F-H-K-H) monitored, the present data showed that two isoforms of eIF3E (Os07g12110 and Os07g07250) and one isoform of eIF3F (Os05g01450) are restricted to the embryo. In Arabidopsis, the eIF3F is a key regulator of embryo development particularly in actively developing tissues (Xia et al., 2010). In addition, it interacts with eIF3E suggesting that the observed rice eIF3E/3F proteins would play significant roles during embryonic or post-embryonic cellular processes. Upon formation, the multifactor complex (MFC) associates with the 40S ribosome thereby establishing the 43S pre-initiation complex (Figure 6D).

In parallel, the mRNA cap is recognized by the eIF4F complex composed of both a eIF4E and a eIF4G protein family (Roy and von Arnim, 2013). In plants, a very important feature is the presence of isoforms of eIF4E and eIF4G named eIF(iso)4E and eIF(iso)4G. These different isoforms participate to the mRNA translational selectivity (Mayberry et al., 2009; Martinez-Silva et al., 2012). In our data, we found that the eIF4E and eIF4(iso)4E subunits, responsible for recognition of the mRNA cap, were restricted to the rice embryo and present in relatively similar abundances (Table S3F, Figure 6D). The 43S pre-initiation complex subsequently associates to the eIF4 complex and binds to the mRNA. Translation initiation factors eIF4A with ATP-dependent helicase activity unwind mRNA 5'UTR secondary structures. The eIF4A-1 and eIF4A-3 were detected in our proteomic data (Table S3F, Figure 6D). The eIF4A-1 translation initiation factor is present in both endosperm and embryo, but, in contrast to almost all translation initiation factors, this protein is significantly and strongly up-accumulated in the endosperm (Table S3F, Figure 6D). The eIF4A-3 protein was strictly observed in the embryo suggesting that embryo and endosperm use different eIF4A helicases during mRNA translation initiation. The translation initiation factor eIF6-2 interacts with RACK1, a negative regulator of ABA response and positive regulator of GA signaling (Guo et al., 2011; Fennell et al., 2012). Especially relevant in seed biology, it was demonstrated that ABA inhibited RACK1 and eIF6 gene expressions (Guo et al., 2011). In Arabidopsis, three homologs of the mammalian RACK1, namely RACK1A, RACK1B and RACK1C were characterized. In our proteomic description of rice seeds, it was interesting to observe the presence of two RACK1 proteins with OsRACK1A detected in both tissues and OsRACK1B only detected in the embryo along with the embryo-specific eIF6-2 (Table S3C, Figure 6D). Recently, the *OsRACK1A* gene has been shown to positively



regulate rice seed germination through promotion of ABA catabolism and  $H_2O_2$  synthesis (Zhang et al., 2014). Thus, in addition to RACK1A, RACK1B could also play a major role in the embryo during seed germination. RACK1A and B could also link ABA and GA signaling with mRNA translation. We only detected the eIF6-2 protein in the embryo suggesting that RACK1A/B regulations of mRNA translation only apply to the embryo and not to the endosperm. Together with the absence of the two cap recognition eIF4E proteins in the endosperm, it further confirms that embryo and endosperm may have contrasted qualitative mRNA translation regulations. Very recent evidences showed that eIFiso4G1 translation initiation factor has a role in the fatty acid profile of Arabidopsis developing seeds through the balance of plastid and nucleus-encoded mRNAs involved in fatty acid biosynthesis (Li et al., 2017). Moreover, this role of eIFisoG1 is not compensated by eIFiso4G2 suggesting a very specific effect. Further work should distinguish the translational machinery at the tissue level and its consequences on seed metabolism.

## CONCLUDING REMARKS

Starch and SSPs were long associated with the endosperm storage function. Thus, it was remarkable to pinpoint, in our proteomic dataset, the presence of glutelins and starch biosynthesis enzymes at non-negligible level also in the embryo (Figures 4, 7). These results refine and expand previous proteomic results on whole developing rice seeds (Koller et al., 2002). Altogether, this would also support the supernumerary embryo hypothesis with both ancestral tissues being equipped with different molecular apparatus before divergence. Classically, the inner SE has been seen as a dead storage tissue since the central parts of the endosperm undergo PCD (Young and Gallie, 2000). Yet, we could show that, as expected, the embryoless endosperm (aleurone and SE) showed an important translational activity however, the embryo display a higher protein synthesis (Figures 6B,C). While the functional consequences of this mRNA translational activity remain to be established, it is clear that the seed endosperm has emerging new roles regarding the control of seed germination and environmental adaptation (Yan et al., 2014; Bassel, 2016). New determinants of agricultural seed quality both in monocots and dicots crops will undoubtedly benefit from tissue-specific combined “multi-omics.” The genetic and tissue heterogeneity of the mature seed is a considerable challenge to seed biologists. In addition, the seed definitely constitute a fascinating plant organ in which post-transcriptional regulations and translational selectivity fine-tune the biological processes that are spatially and temporally regulated within a few hours. A renewed vision of seed biology by integrative systems biology would certainly dig out meaningful new genetic determinants of seed quality.

## AUTHOR CONTRIBUTIONS

MG, DH, IL, and LR designed and performed the experimental work. GCl completed the metabolome analyses. SB and SH

realized the transcriptomic analysis while BV gathered the proteomic data. GC, BC were involved in the preparation of samples. FG, JT, and EA helped with R statistics. HM and BG provided help with cytological observations. MG and LR wrote the manuscript. All authors read and approved the final manuscript.

## FUNDING

This work was supported by the French Ministry of Industry (FUI, NUTRICE agreement # 092906334).

## ACKNOWLEDGMENTS

We are thankful to Loïc Fontaine and Jean-Benoît Morel (UMR BGPI, CIRAD, Montpellier, France) for supply of rice Nipponbare seeds. We acknowledge Dr Francisco Cubillos Riffo (Laboratorio de Microbiología y Biotecnología Aplicada, Centro de Estudios en Ciencia y Tecnología de los Alimentos, Santiago, Chile) for help with statistical analysis. Thanks to Olivier Langella and Thierry Balliau, from PAPPSO (Plateforme d'Analyse Protéomique de Paris Sud-Ouest), for making the proteomic data available on PROTEICdb. The author gratefully acknowledges the support of K. C. Wong Education Foundation, Hong Kong for the visit of DH at IJPB. Thanks to the Saclay Plant Sciences (SPS) LabEx supporting IJPB, IPS2 and GQE-Le Moulon (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.2017.01984/full#supplementary-material>

**Figure S1 |** Protein SDS-PAGE gels used for label-free shotgun proteomics. (A,B) Proteins were extracted, quantified by the Bradford assay and 25 µg of embryo (A) or endosperm (B) proteins were subjected to SDS-PAGE (10% acrylamide) and Coomassie staining. (C) Protein quantification in each lane according to the Multi Gauge 3.0 software. Areas quantified are shown on the gels.

**Figure S2 |** Transcriptome (A) and proteome (B) analysis workflows applied in this study.

**Figure S3 |** Gene Singular Enrichment Analysis of genes preferentially expressed in the endosperm. From the list of the differentially expressed genes, we extracted 787 probe sets with a preferred expression in the endosperm ( $\log_2$  ratio  $> +1.4$ ,  $p < 0.01$ ). The resulting list of 787 probe sets was then submitted to the AgriGO Gene Singular Enrichment Analysis (Hypergeometric test corrected by Yekutieli False Discovery Rate correction, Affymetrix Rice Genome Array set as background,  $p < 0.05$ ) to detect enriched Biological Process (A), Cellular Compartment (B), and Molecular Function (C) GO categories. 722 probes were classified. Light yellow, dark yellow and red boxes indicates GO terms significant enrichment at  $p < 0.05$ ,  $< 0.01$ , and  $< 0.001$  respectively.

**Figure S4 |** Gene Singular Enrichment Analysis of genes preferentially expressed in the embryo. From the list of the differentially expressed genes, we extracted 1,921 probe sets with a preferred expression in the embryo ( $\log_2$  ratio  $< -1$ ,  $p < 0.01$ ). The resulting list of 1,921 probe sets was then submitted to the AgriGO Gene Singular Enrichment Analysis (Hypergeometric test corrected by Yekutieli False Discovery Rate correction, Affymetrix Rice Genome Array set as background,  $p < 0.05$ ) to detect enriched Biological Process (A), Cellular Compartment (B), and Molecular Function (C) GO categories. 1,712 probes were

classified. Light yellow, dark yellow and red boxes indicates GO terms significant enrichment at  $p < 0.05$ ,  $< 0.01$ , and  $< 0.001$  respectively.

**Figure S5 |** Embryo and endosperm peptide and protein distributions. **(A–D)** Distributions of the gene-specific peptide and protein abundances from **(A,C)** embryo and **(B,D)** endosperm on a  $\log_2$ -transformed axis. **(E,F)** Quantile-quantile comparison of embryo and endosperm **(E)** peptide and **(F)** protein abundances ( $\log_2$ -transformed). **(G)** Distribution of the  $\log_2$  ratios (endosperm vs. embryo) for the 673 common proteins. The comparison with the theoretical normal law (red line with a mean equal to the estimated mean i.e.,  $-1.89$  and standard deviation  $2.89$ ) is shown.

**Figure S6 |** Gene Singular Enrichment Analysis of endosperm-specific or endosperm-favored proteins. The list of the 113 endosperm-specific and 76 endosperm-favored ( $\log_2$  ratio  $> 1.7$ ) was subjected to a Gene Singular Enrichment Analysis tool (hypergeometric test corrected by Yekutieli False

Discovery rate, whole transcriptome set as background,  $p < 0.05$ ) to detect enriched endosperm protein Biological Process **(A)**, Cellular Compartment **(B)**, and Molecular Function **(C)** GO categories. 164 proteins were classified. Light yellow, dark yellow and red boxes indicates GO terms significant enrichment at  $p < 0.05$ ,  $< 0.01$ , and  $< 0.001$  respectively.

**Figure S7 |** Gene Singular Enrichment Analysis of embryo-specific or embryo-favored proteins. The list of the 1,426 embryo-specific and 267 embryo-favored ( $\log_2$  ratio  $< -2.9$ ) was subjected to a Gene Singular Enrichment Analysis tool (hypergeometric test corrected by Yekutieli False Discovery rate, whole transcriptome set as background,  $p < 0.05$ ) to detect enriched embryo protein Biological Process **(A)**, Cellular Compartment **(B)**, and Molecular Function **(C)** GO categories. 1,627 proteins were classified. Light yellow and red boxes indicates GO terms significant enrichment at  $p < 0.05$ ,  $< 0.01$ , and  $< 0.001$  respectively.

## REFERENCES

- Alpert, P. (2005). The limits and frontiers of desiccation-tolerant life. *Integr. Comp. Biol.* 45, 685–695. doi: 10.1093/icb/45.5.685
- Arc, E., Galland, M., Cuff, G., Godin, B., Lounifi, I., Job, D., et al. (2011). Reboot the system thanks to protein post-translational modifications and proteome diversity: how quiescent seeds restart their metabolism to prepare seedling establishment. *Proteomics* 11, 1606–1618. doi: 10.1002/pmic.201000641
- Avila-Ospina, L., Clément, G., and Masclaux-Daubresse, C. (2017). Metabolite profiling for leaf senescence in barley reveals decreases in amino acids and glycolysis intermediates. *Agronomy* 7:15. doi: 10.3390/agronomy7010015
- Bailly, C., El-Maarouf-Bouteau, H., and Corbineau, F. (2008). From intracellular signaling networks to cell death: the dual role of reactive oxygen species in seed physiology. *C. R. Biol.* 331, 806–814. doi: 10.1016/j.crv.2008.07.022
- Bao, J., Jin, L., Xiao, P., Shen, S., Sun, M., and Corke, H. (2008). Starch physicochemical properties and their associations with microsatellite alleles of starch-synthesizing genes in a rice RIL population. *J. Agric. Food Chem.* 56, 1589–1594. doi: 10.1021/jf073128+
- Barthole, G., Lepiniec, L., Rogowsky, P. M., and Baud, S. (2012). Controlling lipid accumulation in cereal grains. *Plant Sci.* 185–186, 33–39. doi: 10.1016/j.plantsci.2011.09.002
- Bassel, G. W. (2016). To Grow or not to Grow? *Trends Plant Sci.* 21, 498–505. doi: 10.1016/j.tplants.2016.02.001
- Beckles, D. M., Smith, A. M., and ap Rees, T. (2001). A cytosolic ADP-glucose pyrophosphorylase is a feature of graminaceous endosperms, but not of other starch-storing organs. *Plant Physiol.* 125, 818–827. doi: 10.1104/pp.125.2.818
- Belmonte, M. F., Kirkbride, R. C., Stone, S. L., Pelletier, J. M., Bui, A. Q., Yeung, E. C., et al. (2013). Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. *Proc. Natl. Acad. Sci. U.S.A.* 110, E435–444. doi: 10.1073/pnas.1222061110
- Bentsink, L., Alonso-Blanco, C., Vreugdenhil, D., Tesnier, K., Groot, S. P., and Koornneef, M. (2000). Genetic analysis of seed-soluble oligosaccharides in relation to seed storability of Arabidopsis. *Plant Physiol.* 124, 1595–1604. doi: 10.1104/pp.124.4.1595
- Bethke, P. C., Libourel, I. G., Aoyama, N., Chung, Y. Y., Still, D. W., and Jones, R. L. (2007). The Arabidopsis aleurone layer responds to nitric oxide, gibberellin, and abscisic acid and is sufficient and necessary for seed dormancy. *Plant Physiol.* 143, 1173–1188. doi: 10.1104/pp.106.093435
- Boca, S., Koestler, F., Ksas, B., Chevalier, A., Leymarie, J., Fekete, A., et al. (2013). Arabidopsis lipocalins AtCHL and AtTIL have distinct but overlapping functions essential for lipid protection and seed longevity. *Plant Cell Environ.* 37, 368–381. doi: 10.1111/pce.12159
- Bonhomme, L., Valot, B., Tardieu, F., and Zivy, M. (2012). Phosphoproteome dynamics upon changes in plant water status reveal early events associated with rapid growth adjustment in maize leaves. *Mol. Cell. Proteomics* 11, 957–972. doi: 10.1074/mcp.M111.015867
- Boudet, J., Buitink, J., Hoekstra, F. A., Rogniaux, H., Larre, C., Satour, P., et al. (2006). Comparative analysis of the heat stable proteome of radicles of *Medicago truncatula* seeds during germination identifies late embryogenesis abundant proteins associated with desiccation tolerance. *Plant Physiol.* 140, 1418–1436. doi: 10.1104/pp.105.074039
- 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
- Buitink, J., and Leprince, O. (2004). Glass formation in plant anhydrobiotes: survival in the dry state. *Cryobiology* 48, 215–228. doi: 10.1016/j.cryobiol.2004.02.011
- Buitink, J., and Leprince, O. (2008). Intracellular glasses and seed survival in the dry state. *C. R. Biol.* 331, 788–795. doi: 10.1016/j.crv.2008.08.002
- Charron, J. B., Ouellet, F., Pelletier, M., Danyluk, J., Chauve, C., and Sarhan, F. (2005). Identification, expression, and evolutionary analyses of plant lipocalins. *Plant Physiol.* 139, 2017–2028. doi: 10.1104/pp.105.070466
- Chatelain, E., Hundertmark, M., Leprince, O., Le Gall, S., Satour, P., Deligny-Penninck, S., et al. (2012). Temporal profiling of the heat-stable proteome during late maturation of *Medicago truncatula* seeds identifies a restricted subset of late embryogenesis abundant proteins associated with longevity. *Plant Cell Environ.* 35, 1440–1455. doi: 10.1111/j.1365-3040.2012.02501.x
- Châtelain, E., Satour, P., Laugier, E., Ly Vu, B., Payet, N., Rey, P., et al. (2013). Evidence for participation of the methionine sulfoxide reductase repair system in plant seed longevity. *Proc. Natl. Acad. Sci. U.S.A.* 110, 3633–3638. doi: 10.1073/pnas.1220589110
- Craig, R., and Beavis, R. C. (2004). TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* 20, 1466–1467. doi: 10.1093/bioinformatics/bth092
- Chu, P., Chen, H., Zhou, Y., Li, Y., Ding, Y., Jiang, L., et al. (2012). Proteomic and functional analyses of *Nelumbo nucifera* annexins involved in seed thermotolerance and germination vigor. *Planta* 235, 1271–1288. doi: 10.1007/s00425-011-1573-y
- Clark, G. B., Morgan, R. O., Fernandez, M. P., and Roux, S. J. (2012). Evolutionary adaptation of plant annexins has diversified their molecular structures, interactions and functional roles. *New Phytol.* 196, 695–712. doi: 10.1111/j.1469-8137.2012.04308.x
- Clarke, N. A., Wilkinson, M. C., and Laidman, D. L. (1983). “Lipids in cereal technology,” in *Lipids in Cereal Technology*, ed P. J. Barnes (London: London Academic Press), 57–92.
- De Giorgi, J., Piskurewicz, U., Loubéry, S., Utz-Pugin, A., Bailly, C., Mène-Saffrané, L., et al. (2015). An endosperm-associated cuticle is required for Arabidopsis seed viability, dormancy and early control of germination. *PLoS Genet.* 11:e1005708. doi: 10.1371/journal.pgen.1005708
- Devaiah, S. P., Pan, X., Hong, Y., Roth, M., Welti, R., and Wang, X. (2007). Enhancing seed quality and viability by suppressing phospholipase D in Arabidopsis. *Plant J.* 50, 950–957. doi: 10.1111/j.1365-313X.2007.03103.x
- Du, Z., Zhou, X., Ling, Y., Zhang, Z., and Su, Z. (2010). agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acid Res.* 38, W64–Q70. doi: 10.1093/nar/gkq310
- Ehrenshaft, M., and Brambl, R. (1990). Respiration and mitochondrial biogenesis in germinating embryos of maize. *Plant Physiol.* 93, 295–304. doi: 10.1104/pp.93.1.295

- Ellis, R. H., and Hong, T. D. (1994). Desiccation tolerance and potential longevity of developing seeds of Rice (*Oryza sativa* L.). *Ann. Bot.* 73, 501–506. doi: 10.1006/anbo.1994.1062
- Farrant, J. M., and Moore, J. P. (2011). Programming desiccation-tolerance: from plants to seeds to resurrection plants. *Curr. Opin. Plant Biol.* 14, 340–345. doi: 10.1016/j.pbi.2011.03.018
- Fennell, H., Olawin, A., Mizanur, R. M., Izumori, K., Chen, J. G., and Ullah, H. (2012). Arabidopsis scaffold protein RACK1A modulates rare sugar D-allose regulated gibberellin signaling. *Plant Signal. Behav.* 7, 1407–1410. doi: 10.4161/psb.21995
- Fernandez, O., Bethencourt, L., Quero, A., Sangwan, R. S., and Clement, C. (2010). Trehalose and plant stress responses: friend or foe? *Trends Plant Sci.* 15, 409–417. doi: 10.1016/j.tplants.2010.04.004
- Fiehn, O., Wohlgemuth, G., Scholz, M., Kind, T., Lee, D. Y., Lu, Y., et al. (2008). Quality control for plant metabolomics: reporting MSI-compliant studies. *Plant J.* 53, 691–704. doi: 10.1111/j.1365-3113X.2007.03387.x
- Fincher, G. B. (1989). Molecular and cellular biology associated with endosperm mobilization in germinating cereal grains. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 40, 305–346. doi: 10.1146/annurev.pp.40.060189.001513
- Folsom, J. J., Begcy, K., Hao, X., Wang, D., and Walia, H. (2014). Rice FIE1 regulates seed size under heat stress by controlling early endosperm development. *Plant Physiol.* 165, 238–248. doi: 10.1104/pp.113.232413
- Friedman, W. (1998). The evolution of double fertilization and endosperm: an “historical” perspective. *Sex. Plant Reprod.* 11, 6–16. doi: 10.1007/s004970050114
- Galland, M., Boutet-Mercey, S., Lounifi, I., Godin, B., Balzergue, S., Grandjean, O., et al. (2014a). Compartmentation and dynamics of flavone metabolism in dry and germinated rice seeds. *Plant Cell Physiol.* 55, 1646–1659. doi: 10.1093/pcp/pcu095
- Galland, M., Huguet, R., Arc, E., Cueff, G., Job, D., and Rajjou, L. (2014b). Dynamic proteomics emphasizes the importance of selective mRNA translation and protein turnover during Arabidopsis seed germination. *Mol. Cell. Proteomics* 13, 252–268. doi: 10.1074/mcp.M113.032227
- Galland, M., and Rajjou, L. (2015). Regulation of mRNA translation controls seed germination and is critical for seedling vigor. *Front. Plant Sci.* 6:284. doi: 10.3389/fpls.2015.00284
- Gallardo, K., Firnhaber, C., Zuber, H., Hericher, D., Belghazi, M., Henry, C., et al. (2007). A combined proteome and transcriptome analysis of developing *Medicago truncatula* seeds: evidence for metabolic specialization of maternal and filial tissues. *Mol. Cell. Proteomics* 6, 2165–2179. doi: 10.1074/mcp.M700171-MCP200
- Greenspan, P., Mayer, E. P., and Fowler, S. D. (1985). Nile red: a selective fluorescent stain for intracellular lipid droplets. *J. Cell Biol.* 100, 965–973. doi: 10.1083/jcb.100.3.965
- Grzyb, J., Latowski, D., and Strzalka, K. (2006). Lipocalins—a family portrait. *J. Plant Physiol.* 163, 895–915. doi: 10.1016/j.jplph.2005.12.007
- Guglielminetti, L., Yamaguchi, J., Perata, P., and Alpi, A. (1995). Amylolytic activities in cereal seeds under aerobic and anaerobic conditions. *Plant Physiol.* 109, 1069–1076. doi: 10.1104/pp.109.3.1069
- Guo, J., Wang, S., Valerius, O., Hall, H., Zeng, Q., Li, J. F., et al. (2011). Involvement of Arabidopsis RACK1 in protein translation and its regulation by abscisic acid. *Plant Physiol.* 155, 370–383. doi: 10.1104/pp.110.160663
- Guo, X., Wu, Y., Wang, Y., Chen, Y., and Chu, C. (2009). OsMSRA4.1 and OsMSRB1.1, two rice plastidial methionine sulfoxide reductases, are involved in abiotic stress responses. *Planta* 230, 227–238. doi: 10.1007/s00425-009-0934-2
- Hajdúch, M., Hearne, L. B., Miernyk, J. A., Casteel, J. E., Joshi, T., Agrawal, G. K., et al. (2010). Systems analysis of seed filling in Arabidopsis: using general linear modeling to assess concordance of transcript and protein expression. *Plant Physiol.* 152, 2078–2087. doi: 10.1104/pp.109.152413
- Han, C., He, D., Li, M., and Yang, P. (2014). In-depth proteomic analysis of rice embryo reveals its important roles in seed germination. *Plant Cell Physiol.* 55, 1826–1847. doi: 10.1093/pcp/pcu114
- Hand, S. C., Menze, M. A., Toner, M., Boswell, L., and Moore, D. (2011). LEA proteins during water stress: not just for plants anymore. *Annu. Rev. Physiol.* 73, 115–134. doi: 10.1146/annurev-physiol-012110-142203
- Hay, F. R., and Whitehouse, K. J. (2017). Rethinking the approach to viability monitoring in seed genebanks. *Conserv. Physiol.* 5:cox009. doi: 10.1093/conphys/cox009
- Hinch, D. K., and Thalhammer, A. (2012). LEA proteins: IDPs with versatile functions in cellular dehydration tolerance. *Biochem. Soc. Trans.* 40, 1000–1003. doi: 10.1042/BST20120109
- Hoekstra, F. A., Golovina, E. A., and Buitink, J. (2001). Mechanisms of plant desiccation tolerance. *Trends Plant Sci.* 6, 431–438. doi: 10.1016/S1360-1385(01)02052-0
- Howell, K. A., Millar, A. H., and Whelan, J. (2006). Ordered assembly of mitochondria during rice germination begins with pro-mitochondrial structures rich in components of the protein import apparatus. *Plant Mol. Biol.* 60, 201–223. doi: 10.1007/s11103-005-3688-7
- Howell, K. A., Narsai, R., Carroll, A., Ivanova, A., Lohse, M., Usadel, B., et al. (2009). Mapping metabolic and transcript temporal switches during germination in rice highlights specific transcription factors and the role of RNA instability in the germination process. *Plant Physiol.* 149, 961–980. doi: 10.1104/pp.108.129874
- Huang, J., Cai, M., Long, Q., Liu, L., Lin, Q., Jiang, L., et al. (2014). OsLOX2, a rice type I lipoxygenase, confers opposite effects on seed germination and longevity. *Transgenic Res.* 23, 643–655. doi: 10.1007/s11248-014-9803-2
- Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., et al. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* 12, 115–121. doi: 10.1038/nmeth.3252
- Hundertmark, M., Buitink, J., Leprince, O., and Hinch, D. K. (2011). The reduction of seed-specific dehydrins reduces seed longevity in Arabidopsis thaliana. *Seed Sci. Res.* 21, 165–173. doi: 10.1017/S0960258511000079
- Ida, S., Masaki, Y., and Morita, Y. (1983). The isolation of multiple forms and product specificity of Rice Lipoxygenase. *Agric. Biol. Chem.* 47, 637–641.
- International Rice Genome Sequencing Project (2005). The map-based sequence of the rice genome. *Nature* 436, 793–800. doi: 10.1038/nature03895
- Jeon, J. S., Ryoo, N., Hahn, T. R., Walia, H., and Nakamura, Y. (2010). Starch biosynthesis in cereal endosperm. *Plant Physiol. Biochem.* 48, 383–392. doi: 10.1016/j.plaphy.2010.03.006
- Jiao, Y., Tausta, S. L., Gandotra, N., Sun, N., Liu, T., Clay, N. K., et al. (2009). A transcriptome atlas of rice cell types uncovers cellular, functional and developmental hierarchies. *Nat. Genet.* 41, 258–263. doi: 10.1038/ng.282
- Jolivet, P., Boulard, C., Bellamy, A., Larre, C., Barre, M., Rogniaux, H., et al. (2009). Protein composition of oil bodies from mature *Brassica napus* seeds. *Proteomics* 9, 3268–3284. doi: 10.1002/pmic.200800449
- Joosen, R. V. L., Ligterink, W., Dekkers, B. J., and Hilhorst, H. W. M. (2011). Visualization of molecular processes associated with seed dormancy and germination using MapMan. *Seed Sci. Res.* 21, 143–152. doi: 10.1017/S0960258510000449
- Jung, K. H., An, G., and Ronald, P. C. (2008). Towards a better bowl of rice: assigning function to tens of thousands of rice genes. *Nat. Rev. Genet.* 9, 91–101. doi: 10.1038/nrm2343
- Kaur, H., Petla, B. P., Kamble, N. U., Singh, A., Rao, V., Salvi, P., et al. (2015). Differentially expressed seed aging responsive heat shock protein OsHSP18.2 implicates in seed vigor, longevity and improves germination and seedling establishment under abiotic stress. *Front. Plant Sci.* 6:713. doi: 10.3389/fpls.2015.00713
- Kaur, H., Petla, B. P., and Majee, M. (2016). “Small heat shock proteins: roles in development, desiccation tolerance and seed longevity,” in *Heat Shock Proteins and Plants*, eds A. A. A. Asea, P. Kaur, and S. Calderwood (New York, NY: Springer International Publishing), 3–18.
- Kawahara, Y., de la Bastide, M., Hamilton, J. P., Kanamori, H., McCombie, W. R., Ouyang, S., et al. (2013). Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* 6:4. doi: 10.1186/1939-8433-6-4
- Kawakatsu, T., Yamamoto, M. P., Hirose, S., Yano, M., and Takaiwa, F. (2008). Characterization of a new rice glutelin gene GluD-1 expressed in the starchy endosperm. *J. Exp. Bot.* 59, 4233–4245. doi: 10.1093/jxb/ern265
- Kim, S. T., Wang, Y., Kang, S. Y., Kim, S. G., Rakwal, R., Kim, Y. C., et al. (2009). Developing rice embryo proteomics reveals essential role for embryonic proteins in regulation of seed germination. *J. Proteome Res.* 8, 3598–3605. doi: 10.1021/pr900358s



- Koller, A., Washburn, M. P., Lange, B. M., Andon, N. L., Deciu, C., Haynes, P. A., et al. (2002). Proteomic survey of metabolic pathways in rice. *Proc. Natl. Acad. Sci. U.S.A.* 99, 11969–11974. doi: 10.1073/pnas.172183199
- Kumamaru, T., Uemura, Y., Inoue, Y., Takemoto, Y., Siddiqui, S. U., Ogawa, M., et al. (2010). Vacuolar processing enzyme plays an essential role in the crystalline structure of glutelin in rice seed. *Plant Cell Physiol.* 51, 38–46. doi: 10.1093/pcp/pcp165
- Lafon-Placette, C., and Kohler, C. (2014). Embryo and endosperm, partners in seed development. *Curr. Opin. Plant Biol.* 17, 64–69. doi: 10.1016/j.pbi.2013.11.008
- Lawlor, K. E., and Vince, J. E. (2014). Ambiguities in NLRP3 inflammasome regulation: is there a role for mitochondria? *Biochim. Biophys. Acta* 1840, 1433–1440. doi: 10.1016/j.bbagen.2013.08.014
- Layat, E., Leymarie, J., El-Maarouf-Bouteau, H., Caius, J., Langlade, N., and Bailly, C. (2014). Translational profiling in dormant and nondormant sunflower (*Helianthus annuus*) seeds highlights post-transcriptional regulation of germination. *New Phytol.* 204, 864–872. doi: 10.1111/nph.13002
- Lee, J., and Koh, H. J. (2011). A label-free quantitative shotgun proteomics analysis of rice grain development. *Proteome Sci.* 9:61. doi: 10.1186/1477-5956-9-61
- Leonova, S., Grimberg, A., Marttila, S., Szymne, S., and Carlsson, A. S. (2010). Mobilization of lipid reserves during germination of oat (*Avena sativa* L.), a cereal rich in endosperm oil. *J. Exp. Bot.* 61, 3089–3099. doi: 10.1093/jxb/erq141
- Li, G., Lin, F., and Xue, H. W. (2007). Genome-wide analysis of the phospholipase D family in *Oryza sativa* and functional characterization of PLD beta 1 in seed germination. *Cell Res.* 17, 881–894. doi: 10.1038/cr.2007.77
- Li, Q., Shen, W., Zheng, Q., Tan, Y., Gao, J., Shen, J., et al. (2017). Effects of eIFiso4G1 mutation on seed oil biosynthesis. *Plant J.* 90, 966–978. doi: 10.1111/tpj.13522
- Logan, D. C., Millar, A. H., Sweetlove, L. J., Hill, S. A., and Leaver, C. J. (2001). Mitochondrial biogenesis during germination in maize embryos. *Plant Physiol.* 125, 662–672. doi: 10.1104/pp.125.2.662
- Lu, Y., Wu, Y. R., and Han, B. (2005). Anaerobic induction of isocitrate lyase and malate synthase in submerged rice seedlings indicates the important metabolic role of the glyoxylate cycle. *Acta Biochim. Biophys. Sin.* 37, 406–414. doi: 10.1111/j.1745-7270.2005.00060.x
- Martínez-Silva, A. V., Aguirre-Martínez, C., Flores-Tinoco, C. E., Alejandri-Ramírez, N. D., and Dinkova, T. D. (2012). Translation initiation factor AtelF(iso)4E is involved in selective mRNA translation in *Arabidopsis thaliana* seedlings. *PLoS ONE* 7:e31606. doi: 10.1371/journal.pone.0031606
- Mayberry, L. K., Allen, M. L., Dennis, M. D., and Browning, K. S. (2009). Evidence for variation in the optimal translation initiation complex: plant eIF4B, eIF4E, and eIF(iso)4F differentially promote translation of mRNAs. *Plant Physiol.* 150, 1844–1854. doi: 10.1104/pp.109.138438
- Moskovitz, J. (2005). Methionine sulfoxide reductases: ubiquitous enzymes involved in antioxidant defense, protein regulation, and prevention of aging-associated diseases. *Biochim. Biophys. Acta* 1703, 213–219. doi: 10.1016/j.bbapap.2004.09.003
- Mudgett, M. B., and Clarke, S. (1994). Hormonal and environmental responsiveness of a developmentally regulated protein repair L-isoaspartyl methyltransferase in wheat. *J. Biol. Chem.* 269, 25605–25612.
- Murphy, D. J. (1993). Structure, function and biogenesis of storage lipid bodies and oleosins in plants. *Prog. Lipid Res.* 32, 247–280. doi: 10.1016/0163-7827(93)90009-L
- Murphy, D. J., Hernandez-Pinzon, I., and Patel, K. (2001). Role of lipid bodies and lipid-body proteins in seeds and other tissues. *J. Plant Physiol.* 158, 471–478. doi: 10.1078/0176-1617-00359
- Narayana, N. K., Vemanna, R. S., Kodekallu Chandrashekar, B., Rao, H., Vennapusa, A. R., Narasimaha, A., et al. (2017). Aldo-ketoreductase 1 (AKR1) improves seed longevity in tobacco and rice by detoxifying reactive cytotoxic compounds generated during ageing. *Rice* 10:11. doi: 10.1186/s12284-017-0148-3
- Neuberger, T., Sreenivasulu, N., Rokitta, M., Rolletschek, H., Gobel, C., Rutten, T., et al. (2008). Quantitative imaging of oil storage in developing crop seeds. *Plant Biotechnol. J.* 6, 31–45. doi: 10.1111/j.1467-7652.2007.00294.x
- Nguyen, H. C., Hoefgen, R., and Hesse, H. (2012). Improving the nutritive value of rice seeds: elevation of cysteine and methionine contents in rice plants by ectopic expression of a bacterial serine acetyltransferase. *J. Exp. Bot.* 63, 5991–6001. doi: 10.1093/jxb/ers253
- Nguyen, T. P., Cuffe, G., Hegedus, D. D., Rajjou, L., and Bentsink, L. (2015). A role for seed storage proteins in *Arabidopsis* seed longevity. *J. Exp. Bot.* 66, 6399–6413. doi: 10.1093/jxb/erv348
- Nowack, M. K., Ungu, A., Bjerkan, K. N., Grini, P. E., and Schnittger, A. (2010). Reproductive cross-talk: seed development in flowering plants. *Biochem. Soc. Trans.* 38, 604–612. doi: 10.1042/BST0380604
- Ogé, L., Bourdais, G., Bove, J., Collet, B., Godin, B., Granier, F., et al. (2008). Protein repair L-isoaspartyl methyltransferase 1 is involved in both seed longevity and germination vigor in *Arabidopsis*. *Plant Cell* 20, 3022–3037. doi: 10.1105/tpc.108.058479
- Olsen, O. A. (2004). Nuclear endosperm development in cereals and *Arabidopsis thaliana*. *Plant Cell* 16, 214–227. doi: 10.1105/tpc.017111
- Ouyang, S., Zhu, W., Hamilton, J., Lin, H., Campbell, M., Childs, K., et al. (2007). The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic Acids Res.* 35(suppl\_1):D883–D887. doi: 10.1093/nar/gkl976
- Paradiso, A., de Pinto, M. C., Locato, V., and De Gara, L. (2012). Galactone-gamma-lactone-dependent ascorbate biosynthesis alters wheat kernel maturation. *Plant Biol.* 14, 652–658. doi: 10.1111/j.1438-8677.2011.00543.x
- Penfield, S., Li, Y., Gilday, A. D., Graham, S., and Graham, I. A. (2006). *Arabidopsis* ABA INSENSITIVE4 regulates lipid mobilization in the embryo and reveals repression of seed germination by the endosperm. *Plant Cell* 18, 1887–1899. doi: 10.1105/tpc.106.041277
- Petla, B. P., Kamble, N. U., Kumar, M., Verma, P., Ghosh, S., Singh, A., et al. (2016). The PROTEIN L-ISOASPARTYL METHYLTRANSFERASE isoforms differentially accumulate during seed maturation to restrict deleterious isoAsp and reactive oxygen species accumulation and are implicated in seed vigor and longevity. *New Phytol.* 211, 627–645. doi: 10.1111/nph.13923
- Rajjou, L., Belghazi, M., Huguet, R., Robin, C., Moreau, A., Job, C., et al. (2006). Proteomic investigation of the effect of salicylic acid on *Arabidopsis* seed germination and establishment of early defense mechanisms. *Plant Physiol.* 141, 910–923. doi: 10.1104/pp.106.082057
- Rajjou, L., and Debeaujon, I. (2008). Seed longevity: survival and maintenance of high germination ability of dry seeds. *C. R. Biol.* 331, 796–805. doi: 10.1016/j.crv.2008.07.021
- Rajjou, L., Duval, M., Gallardo, K., Catusse, J., Bally, J., Job, C., et al. (2012). Seed germination and vigor. *Annu. Rev. Plant Biol.* 63, 507–533. doi: 10.1146/annurev-arplant-042811-105550
- Rajjou, L., Gallardo, K., Debeaujon, I., Vandekerckhove, J., Job, C., and Job, D. (2004). The effect of alpha-amanitin on the *Arabidopsis* seed proteome highlights the distinct roles of stored and neosynthesized mRNAs during germination. *Plant Physiol.* 134, 1598–1613. doi: 10.1104/pp.103.036293
- Rajjou, L., Lovigny, Y., Groot, S. P., Belghazi, M., Job, C., and Job, D. (2008). Proteome-wide characterization of seed aging in *Arabidopsis*: a comparison between artificial and natural aging protocols. *Plant Physiol.* 148, 620–641. doi: 10.1104/pp.108.123141
- Rolletschek, H., Melkus, G., Grafahrend-Belau, E., Fuchs, J., Heinzel, N., Schreiber, F., et al. (2011). Combined noninvasive imaging and modeling approaches reveal metabolic compartmentation in the barley endosperm. *Plant Cell* 23, 3041–3054. doi: 10.1105/tpc.111.087015
- Rouhier, N., Vieira Dos Santos, C., Tarrago, L., and Rey, P. (2006). Plant methionine sulfoxide reductase A and B multigenic families. *Photosyn. Res.* 89, 247–262. doi: 10.1007/s1120-006-9097-1
- Roy, B., and von Arnim, A. G. (2013). Translational regulation of cytoplasmic mRNAs. *The Arabidopsis Book* 11:e0165. doi: 10.1199/tab.0165
- Sano, N., Masaki, S., Tanabata, T., Yamada, T., Hirasawa, T., Kashiwagi, M., et al. (2013). RNA-binding proteins associated with desiccation during seed development in rice. *Biotechnol. Lett.* 35, 1945–1952. doi: 10.1007/s10529-013-1283-6
- Sano, N., Permana, H., Kumada, R., Shinokaki, Y., Tanabata, T., Yamada, T., et al. (2012). Proteomic analysis of embryonic proteins synthesized from long-lived mRNAs during germination of rice seeds. *Plant Cell Physiol.* 53, 687–698. doi: 10.1093/pcp/pcs024
- Sano, N., Rajjou, L., North, H. M., Debeaujon, I., Marion-Poll, A., and Seo, M. (2016). Staying alive: molecular aspects of seed longevity. *Plant Cell Physiol.* 57, 660–674. doi: 10.1093/pcp/pcv186
- Sarkar, N. K., Kim, Y. K., and Grover, A. (2009). Rice sHsp genes: genomic organization and expression profiling under stress and development. *BMC Genomics* 10:393. doi: 10.1186/1471-2164-10-393



- Sasaki, K., Takeuchi, Y., Miura, K., Yamaguchi, T., Ando, T., Ebitani, T., et al. (2015). Fine mapping of a major quantitative trait locus, qLG-9, that controls seed longevity in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* 128, 769–778. doi: 10.1007/s00122-015-2471-7
- Sattler, S. E., Gilliland, L. U., Magallanes-Lundback, M., Pollard, M., and DellaPenna, D. (2004). Vitamin E is essential for seed longevity and for preventing lipid peroxidation during germination. *Plant Cell* 16, 1419–1432. doi: 10.1105/tpc.021360
- Sengupta, D., Naik, D., and Reddy, A. R. (2015). Plant aldo-keto reductases (AKRs) as multi-tasking soldiers involved in diverse plant metabolic processes and stress defense: a structure-function update. *J. Plant Physiol.* 179, 40–55. doi: 10.1016/j.jplph.2015.03.004
- Shin, J. H., Kim, S. R., and An, G. (2009). Rice aldehyde dehydrogenase7 is needed for seed maturation and viability. *Plant Physiol.* 149, 905–915. doi: 10.1104/pp.108.130716
- Sievers, F., Wilm, A., Dineen, D., 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
- Sreenivasulu, N., Usadel, B., Winter, A., Radchuk, V., Scholz, U., Stein, N., et al. (2008). Barley grain maturation and germination: metabolic pathway and regulatory network commonalities and differences highlighted by new MapMan/PageMan profiling tools. *Plant Physiol.* 146, 1738–1758. doi: 10.1104/pp.107.111781
- Subbarao, K. V., Datta, R., and Sharma, R. (1998). Amylases synthesis in scutellum and aleurone layer of maize seeds. *Phytochemistry* 49, 657–666. doi: 10.1016/S0031-9422(97)00964-3
- Takai, K., and Endo, Y. (2010). The cell-free protein synthesis system from wheat germ. *Methods Mol. Biol.* 607, 23–30. doi: 10.1007/978-1-60327-331-2\_3
- Tarpley, L., Duran, A. L., Kebrom, T. H., and Sumner, L. W. (2005). Biomarker metabolites capturing the metabolite variance present in a rice plant developmental period. *BMC Plant Biol.* 5:8. doi: 10.1186/1471-2229-5-8
- Thapar, N., Kim, A. K., and Clarke, S. (2001). Distinct patterns of expression but similar biochemical properties of protein L-isoaspartyl methyltransferase in higher plants. *Plant Physiol.* 125, 1023–1035. doi: 10.1104/pp.125.2.1023
- Valot, B., Langella, O., Nano, E., and Zivy, M. (2011). MassChroQ: a versatile tool for mass spectrometry quantification. *Proteomics* 11, 3572–3577. doi: 10.1002/pmic.201100120
- Verdier, J., Lalanne, D., Pelletier, S., Torres-Jerez, I., Righetti, K., Bandyopadhyay, K., et al. (2013). A regulatory network-based approach dissects late maturation processes related to the acquisition of desiccation tolerance and longevity of *Medicago truncatula* seeds. *Plant Physiol.* 163, 757–774. doi: 10.1104/pp.113.222380
- Walbot, V., and Evans, M. M. (2003). Unique features of the plant life cycle and their consequences. *Nat. Rev. Genet.* 4, 369–379. doi: 10.1038/nrg1064
- Wang, L., Xie, W., Chen, Y., Tang, W., Yang, J., Ye, R., et al. (2010). A dynamic gene expression atlas covering the entire life cycle of rice. *Plant J.* 61, 752–766. doi: 10.1111/j.1365-3113X.2009.04100.x
- Wang, X.-S., Zhu, H.-B., Jin, G.-L., Liu, H.-L., Wu, W.-R., and Zhu, J. (2007). Genome-scale identification and analysis of LEA genes in rice (*Oryza sativa* L.). *Plant Sci.* 172, 414–420. doi: 10.1016/j.plantsci.2006.10.004
- Wang, Y., Zhu, S., Liu, S., Jiang, L., Chen, L., Ren, Y., et al. (2009). The vacuolar processing enzyme OsVPE1 is required for efficient glutelin processing in rice. *Plant J.* 58, 606–617. doi: 10.1111/j.1365-3113X.2009.03801.x
- Waters, E. R. (2013). The evolution, function, structure, and expression of the plant sHSPs. *J. Exp. Bot.* 64, 391–403. doi: 10.1093/jxb/ers355
- Westengen, O. T., Jeppson, S., and Guarino, L. (2013). Global *ex-situ* crop diversity conservation and the Svalbard Global Seed Vault: assessing the current status. *PLoS ONE* 8:e64146. doi: 10.1371/journal.pone.0064146
- Wu, X., Gong, F., Yang, L., Hu, X., Tai, F., and Wang, W. (2014). Proteomic analysis reveals differential accumulation of small heat shock proteins and late embryogenesis abundant proteins between ABA-deficient mutant vp5 seeds and wild-type Vp5 seeds in maize. *Front. Plant Sci.* 5:801. doi: 10.3389/fpls.2014.00801
- Wyse Jackson, P., and Kennedy, K. (2009). The global strategy for plant conservation: a challenge and opportunity for the international community. *Trends Plant Sci.* 14, 578–580. doi: 10.1016/j.tplants.2009.08.011
- Xia, C., Wang, Y. J., Li, W. Q., Chen, Y. R., Deng, Y., Zhang, X. Q., et al. (2010). The Arabidopsis eukaryotic translation initiation factor 3, subunit F (AtelF3f), is required for pollen germination and embryogenesis. *Plant J.* 63, 189–202. doi: 10.1111/j.1365-3113X.2010.04237.x
- Xu, H., Wei, Y., Zhu, Y., Lian, L., Xie, H., Cai, Q., et al. (2015). Antisense suppression of LOX3 gene expression in rice endosperm enhances seed longevity. *Plant Biotechnol. J.* 13, 526–539. doi: 10.1111/pbi.12277
- Xu, S. B., Li, T., Deng, Z. Y., Chong, K., Xue, Y., and Wang, T. (2008). Dynamic proteomic analysis reveals a switch between central carbon metabolism and alcoholic fermentation in rice filling grains. *Plant Physiol.* 148, 908–925. doi: 10.1104/pp.108.125633
- Xue, L. J., Zhang, J. J., and Xue, H. W. (2012). Genome-wide analysis of the complex transcriptional networks of rice developing seeds. *PLoS ONE* 7:e31081. doi: 10.1371/journal.pone.0031081
- Yan, D., Duermeier, L., Leoveanu, C., and Nambara, E. (2014). The functions of the endosperm during seed germination. *Plant Cell Physiol.* 55, 1521–1533. doi: 10.1093/pcp/pcu089
- Yang, P., Li, X., Wang, X., Chen, H., Chen, F., and Shen, S. (2007). Proteomic analysis of rice (*Oryza sativa*) seeds during germination. *Proteomics* 7, 3358–3368. doi: 10.1002/pmic.200700207
- Ye, N., Zhu, G., Liu, Y., Zhang, A., Li, Y., Liu, R., et al. (2012). Ascorbic acid and reactive oxygen species are involved in the inhibition of seed germination by abscisic acid in rice seeds. *J. Exp. Bot.* 63, 1809–1822. doi: 10.1093/jxb/err336
- Yeh, C. H., Yeh, K. W., Wu, S. H., Chang, P. F., Chen, Y. M., and Lin, C. Y. (1995). A recombinant rice 16.9-kDa heat shock protein can provide thermoprotection *in vitro*. *Plant Cell Physiol.* 36, 1341–1348.
- Young, T. E., and Gallie, D. R. (2000). Programmed cell death during endosperm development. *Plant Mol. Biol.* 44, 283–301. doi: 10.1023/A:1026588408152
- Zhan, J., Thakare, D., Ma, C., Lloyd, A., Nixon, N. M., Arakaki, A. M., et al. (2015). RNA sequencing of laser-capture microdissected compartments of the maize kernel identifies regulatory modules associated with endosperm cell differentiation. *Plant Cell* 27, 513–531. doi: 10.1105/tpc.114.135657
- Zhang, D., Chen, L., Li, D., Lv, B., Chen, Y., Chen, J., et al. (2014). OsRACK1 is involved in abscisic acid- and H<sub>2</sub>O<sub>2</sub>-mediated signaling to regulate seed germination in rice (*Oryza sativa* L.). *PLoS ONE* 9:e97120. doi: 10.1371/journal.pone.0097120
- Zhang, Y. X., Xu, H. H., Liu, S. J., Li, N., Wang, W. Q., Möller, I. M., et al. (2016). Proteomic analysis reveals different involvement of embryo and endosperm proteins during aging of Yliangyou 2 hybrid rice seeds. *Front. Plant Sci.* 7:1394. doi: 10.3389/fpls.2016.01394
- Zhu, C., Cheng, C., Liu, X., and He, J. Y. (2007). Accumulation of soluble sugars related to desiccation tolerance during rice (*Oryza sativa* L.) seed development and maturation. *Seed Sci. Technol.* 35, 649–659. doi: 10.15258/sst.2007.35.3.12

**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.

The reviewer ME and handling Editor declared their shared affiliation.

Received: 24 June 2017; Accepted: 03 November 2017; Published: 22 November 2017

Citation: Galland M, He D, Lounifi I, Arc E, Clément G, Balzergue S, Huguet S, Cuffe G, Godin B, Collet B, Granier F, Morin H, Tran J, Valot B and Rajjou L (2017) An Integrated “Multi-Omics” Comparison of Embryo and Endosperm Tissue-Specific Features and Their Impact on Rice Seed Quality. *Front. Plant Sci.* 8:1984. doi: 10.3389/fpls.2017.01984

This article was submitted to Plant Systems and Synthetic Biology, a section of the journal *Frontiers in Plant Science*

Copyright © 2017 Galland, He, Lounifi, Arc, Clément, Balzergue, Huguet, Cuffe, Godin, Collet, Granier, Morin, Tran, Valot and Rajjou. 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) or licensor 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](http://www.frontiersin.org)

**Contact us:** [info@frontiersin.org](mailto: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