

# Metabolic fingerprinting of *Arabidopsis thaliana* accessions

Mariana Sotelo-Silveira<sup>1,2</sup>, Anne-Laure Chauvin<sup>1</sup>, Nayelli Marsch-Martínez<sup>3</sup>, Robert Winkler<sup>3\*</sup> and Stefan de Folter<sup>1\*</sup>

<sup>1</sup> Unidad de Genómica Avanzada (LANGEBIO), Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (CINVESTAV-IPN), Irapuato, México, <sup>2</sup> Laboratorio de Bioquímica, Departamento de Biología Vegetal, Facultad de Agronomía, Universidad de la República, Montevideo, Uruguay, <sup>3</sup> Department of Biotechnology and Biochemistry, CINVESTAV Unidad Irapuato, Irapuato, Mexico

## OPEN ACCESS

### Edited by:

Hirokazu Tsukaya,  
The University of Tokyo, Japan

### Reviewed by:

Masami Yokota Hirai,  
RIKEN Center for Sustainable  
Resource Science, Japan  
Daniel J. Kliebenstein,  
University of California, Davis, USA

### \*Correspondence:

Robert Winkler,  
Department of Biotechnology and  
Biochemistry, CINVESTAV Unidad  
Irapuato, Km. 9.6 Libramiento Norte  
Carr. Irapuato-León, 36821 Irapuato,  
México  
robert.winkler@ira.cinvestav.mx;  
Stefan de Folter,  
Unidad de Genómica Avanzada  
(LANGEBIO), Centro de Investigación  
y de Estudios Avanzados del Instituto  
Politécnico Nacional  
(CINVESTAV-IPN), Km. 9.6  
Libramiento Norte, Carretera  
Irapuato-León, CP 36821 Irapuato,  
Guanajuato, Mexico  
sdfolter@langebio.cinvestav.mx

### Specialty section:

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

Received: 21 January 2015

Accepted: 08 May 2015

Published: 27 May 2015

### Citation:

Sotelo-Silveira M, Chauvin A-L,  
Marsch-Martínez N, Winkler R and de  
Folter S (2015) Metabolic  
fingerprinting of *Arabidopsis thaliana*  
accessions. *Front. Plant Sci.* 6:365.  
doi: 10.3389/fpls.2015.00365

In the post-genomic era much effort has been put on the discovery of gene function using functional genomics. Despite the advances achieved by these technologies in the understanding of gene function at the genomic and proteomic level, there is still a big genotype-phenotype gap. Metabolic profiling has been used to analyze organisms that have already been characterized genetically. However, there is a small number of studies comparing the metabolic profile of different tissues of distinct accessions. Here, we report the detection of over 14,000 and 17,000 features in inflorescences and leaves, respectively, in two widely used *Arabidopsis thaliana* accessions. A predictive Random Forest Model was developed, which was able to reliably classify tissue type and accession of samples based on LC-MS profile. Thereby we demonstrate that the morphological differences among *A. thaliana* accessions are reflected also as distinct metabolic phenotypes within leaves and inflorescences.

**Keywords:** metabolic phenotyping, *Arabidopsis*, accessions, development, metabolites

## Introduction

Biodiversity constitutes a valuable resource for searching genes of interest. Natural variation in *Arabidopsis* has been observed for a variety of traits (Koornneef et al., 2004; Weigel, 2012) like seed size (Alonso-Blanco et al., 1999), light and hormone sensitivity (Maloof et al., 2001), growth rate (Beemster et al., 2002), root growth responses to phosphate starvation (Chevalier et al., 2003), and cold stress responses (Barah et al., 2013), among others. Comparison of whole genomes from *Arabidopsis thaliana* accessions showed that genetic differences exist among them, for instance, over 200 genes found in different accessions are not present in the reference genome Col-0 (Gan et al., 2011; Schneeberger et al., 2011). Furthermore, natural variation has also been studied at the transcriptomic (Gan et al., 2011; Stein and Waters, 2012; Wang et al., 2013) and proteomic (Chevalier et al., 2004) levels.

Metabolomics is adding another dimension to investigate gene function (Fiehn et al., 2000; Saito and Matsuda, 2010). Metabolic analysis methods such as profiling and fingerprinting have evolved from diagnostic tools used to elucidate metabolite accumulation patterns in different tissues and cell compartments of individual plants (Matsuda et al., 2009, 2010, 2011; Krueger et al., 2011; Mintz-Oron et al., 2012) to integrative tools, enhancing the strength of functional genomics in the process of shortening the distance of the genotype-phenotype gap (Fiehn et al., 2000; Taylor et al., 2002; Enot and Draper, 2007; Fernie and Schauer, 2009; García-Flores et al., 2012, 2015; Landesfeind et al., 2014). Recently, the attention in this area has expanded to the study

of natural variation of metabolite levels between individual plants, a strategy that is suggested to provide useful information to improve crop quality (Fernie and Schauer, 2009; Montero-Vargas et al., 2013). In this sense, several studies in Arabidopsis combining metabolomic and QTL analysis showed that metabolite variation between different accessions exists (Keurentjes et al., 2006, 2008; Rowe et al., 2008; Fu et al., 2009; Chan et al., 2010; Joseph et al., 2013, 2014), and highlighted that interactions between transcript and metabolite variation are complex and governed by epistatic interactions (Wentzell et al., 2007; Rowe et al., 2008; Joseph et al., 2013, 2014). Moreover, the metabolic relationship between accessions depends on different factors like tissue, plant age, and environment (Wentzell et al., 2008; Wentzell and Kliebenstein, 2008; Houshyani et al., 2012).

In the present work, we present a metabolite profiling study of *A. thaliana* accessions frequently used in the laboratory: Columbia (Col-0) and Wassilewskija (Ws-3) (Alonso-Blanco and Koornneef, 2000). Col-0 was selected from the original Laibach Landsberg population and is the accession that was sequenced in the Arabidopsis Genome Initiative (Rédei, 1992; AGI, 2000), and Ws-3 is a Russian accession (Laibach, 1951).

We investigated whether a distinct metabolic phenotype in two different tissues could be distinguished besides the morphological and developmental differences observed among the Arabidopsis accessions.

## Material and Methods

### Plant Growth and Plant Material

Col-0 and Ws-3 accessions of Arabidopsis (*A. thaliana*) plants were germinated in soil (3:1:1, peat moss:perlite:vermiculite) in a growth chamber at 22°C under long-day conditions (16 h of light/8 h of dark) and transferred to standard greenhouse conditions (22–27°C, natural light). All plants were grown at the same time under the same environmental conditions.

### Sample Preparation

Fully expanded leaves after flowering, and inflorescences from 10 plants were collected and pooled. Each pool was collected from different plants. Three biological replicas were used for each accession (with exception of Ws-3 leaves with only two biological replicas). Frozen plant material (fully expanded leaves after flowering or whole flowers) was ground in liquid nitrogen. For each 100 mg of fresh tissue, 300 µL of cold acetone was added, and the mixture was vortexed, sonicated for 5 min, and then centrifuged at 16,100 g to separate the crude extract from the tissue, as previously described (Sotelo-Silveira et al., 2013). The supernatant was lyophilized and used for analysis. The lyophilized samples were dissolved in 1000 µL of 100% MeOH and filtered through a 0.22 µm filter before the injection into the chromatographic column. We used for each biological replica two analytical replicas, giving in total 12 inflorescence and 10 leaf samples that were injected (SQLite database; Supplemental Data 3).

## Chromatography

Chromatographic separation was performed on a ACQUITY BEH C-18 column (2.1 × 50 mm i.d., 1.7 µm, Waters, Mexico) using an ACQUITY UPLC system (Waters Corps., Mexico), as previously described (Sotelo-Silveira et al., 2013). The column was maintained at 35°C and eluted with a 30 min gradient. The mobile phase, at a flow rate of 0.2 mL/min, consisted of a starting mixture of solvents A: B (MeOH: H<sub>2</sub>O; 1: 9; A: 100% MeOH; and B: H<sub>2</sub>O + 0.1% formic acid). A decrease of solvent B up to 20% over 15 min was then performed. Solvent B was returned to its initial composition over 1 min and the initial condition was maintained for 15 min in order to equilibrate the column. The volume of sample injected onto the column was 5 µL.

## Mass Spectrometry

The eluent was introduced into the Q-ToF mass spectrometer (LCT Premier™ XE, Waters Corps. Mexico) by electrospray ionization, with capillary and cone voltages set in the positive ion mode to 3100 and 70 V, as previously described (Sotelo-Silveira et al., 2013). The desolvation gas was set to 850 L/h at a temperature of 350°C for the positive mode. The cone gas was set to 10 L/h, and the source temperature to 80°C for the positive mode. Continuum data were acquired from m/z 50–1000 using an accumulation time of 0.2 s per spectrum. All spectra were mass corrected in real-time by reference to leucine enkephalin (2 µg/mL), infused at 5 µL/min through an independent reference electrospray. The resolution of the system was 11,000 for the positive mode.

## Data Analysis

Waters LCT Premier™ XE \*.raw data files were converted to \*.mzML community standard data format using the ProteoWizard (Chambers et al., 2012) and processed with an OpenMS/TOPPAS pipeline (Sturm et al., 2008). A TOPPAS workflow containing the detailed parameters is provided as Supplemental Material (Supplemental Data 1). In short, the LC-MS features of each data set were detected with the FeatureFinderMetabo tool and subsequently merged to create a consensus map. The consensus features were exported to plain text format and manually analyzed using standard text processing and spreadsheet programs.

Only high-quality (HQ) features, which were quantified in all evaluated 12 inflorescence or all 10 leaf samples, respectively, were used for further data analyses. In total 803 such HQ features were found for the inflorescence samples and 561 for the leaf samples. For identifying the HQ features, a metabolite database (DB) for Arabidopsis was created from the KNApSACk database ([http://kanaya.naist.jp/knapsack\\_jsp/top.html](http://kanaya.naist.jp/knapsack_jsp/top.html)) (Afendi et al., 2012) and experimental liquid-chromatograph mass spectrometry (LC-MS) literature data. Automated DB generation and MS data matching were performed using SpiderMass (Winkler, 2015). The SpiderMass Meta-DB for Arabidopsis is provided as Supplemental Data 2. Mass spectrometry data processing was performed on the analysis platform MASSyPup (Winkler, 2014).

Consensus features, HQ features and putative metabolite identifications with their compound classes were integrated

into a SQLite (<https://sqlite.org/>) database, which is available as Supplemental Data 3. For statistical analysis we used the R script “Metabr” (Ernest et al., 2012), which calculates the fold-changes and *p*-values according to Tukey’s Honest Significant Difference (HSD).

The R package and Graphical User Interface (GUI) “Rattle” (Williams, 2009, 2011) was employed to create and evaluate classification models for the metabolic data sets. Due to the characteristics of the data, i.e., relatively few samples, but multiple numeric variables, we decided for a Random Forest model (Williams, 1987, 1988) (Figure 3). For the model training sets, we only considered features present in all 22 LC-MS datasets. To avoid the necessity of imputing values, all features with missing data were omitted. We created a meta-variable “Accession\_Tissue,” which describes the four possible combinations of our experiment and which was used as the target variable. In total, 16 datasets with 460 metabolic input variables were used for the model building. 10,000 decision trees were calculated. The number of selected variables was set to the square root of the number of variables (suggested default), which corresponded to 21 variables. The evaluation of the models was done with five testing datasets that represented the four possible combinations of tissue type and accession. The Rattle project, which contains the final model and the metabolic feature data, is provided as Supplemental Data 4.

## Results

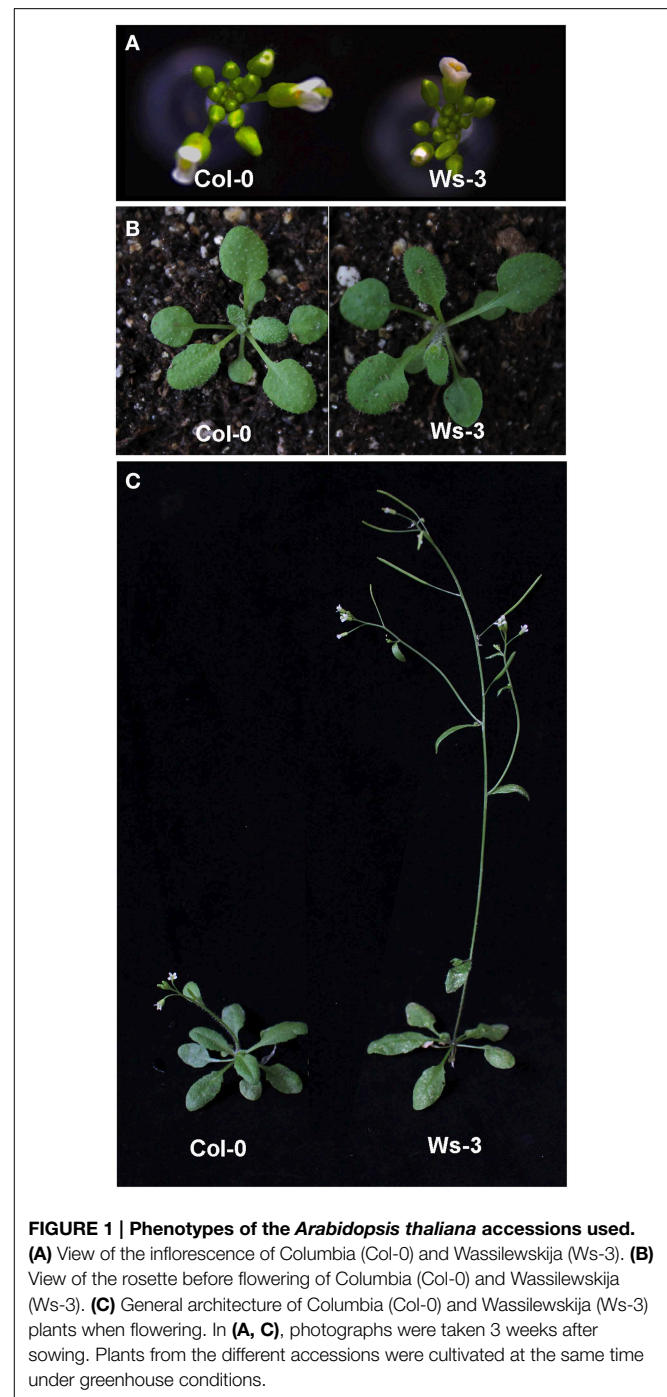
### Selection of Accessions

*A. thaliana* has over 1000 natural accessions that have been collected from around the world (Alonso-Blanco and Koornneef, 2000; Gaut, 2012; Horton et al., 2012). Natural accessions are very variable in terms of shape, development, and physiology (Weigel, 2012). Plants of the commonly used laboratory strains (or accessions), Columbia (Col-0) and Wassilewskija (Ws-3), are distinguishable based on their morphology and development (Figure 1). Particularly, they show differences in rosette leaf development and flowering time. Col-0 plants produce more rosette leaves, have a longer duration of the leaf production period (i.e., they flower later), and have a final rosette leaf area significantly larger than Ws-3 (Massonnet et al., 2010) (Figure 1).

### Distinct Metabolic Phenotypes Were Detected for Each Accession and Tissue

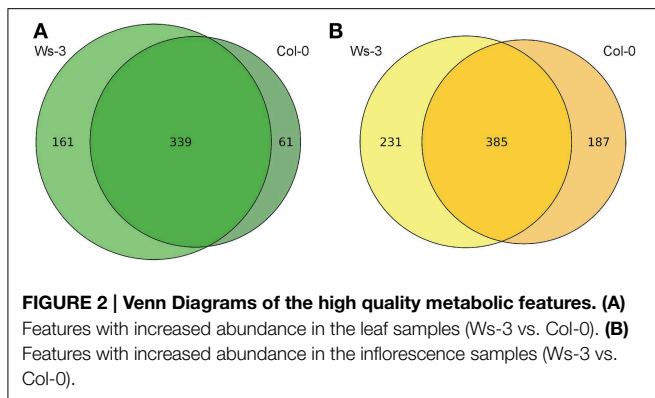
To assess the natural variation in metabolite content among *Arabidopsis* accessions in two different tissues, we performed UPLC-QTOF MS-based untargeted metabolic fingerprinting of crude acetone extracts from leaves and inflorescences collected and pooled from Col-0 and Ws-3. Notably, using an organic solvent favors an extraction toward hydrophobic compounds, which are under-represented in studies using polar solvent mixtures.

The metabolic profiles demonstrate considerable quantitative and qualitative differences between the tissues and accessions. More than 14,000 and 17,000 features from inflorescences and leaves, respectively, were detected in the two accessions



**FIGURE 1 | Phenotypes of the *Arabidopsis thaliana* accessions used. (A)** View of the inflorescence of Columbia (Col-0) and Wassilewskija (Ws-3). **(B)** View of the rosette before flowering of Columbia (Col-0) and Wassilewskija (Ws-3). **(C)** General architecture of Columbia (Col-0) and Wassilewskija (Ws-3) plants when flowering. In **(A, C)**, photographs were taken 3 weeks after sowing. Plants from the different accessions were cultivated at the same time under greenhouse conditions.

(SQLite database; Supplemental Data 3). In total 803 high quality features from inflorescences and 561 from leaves, which were quantified in all evaluated 12 inflorescence or all 10 leaf samples, respectively, were considered for identification through searching in databases of metabolites (Supplemental Data 2). In leaf samples, 222 high quality features presented significant differences ( $p \leq 0.01$ ) and in inflorescences samples 418 high quality features (Figure 2). From these metabolites that presented significant differences we could putatively identify 26 and 36



metabolites in leaf and in inflorescence samples, respectively (Tables 1, 2).

To evaluate the possibility, to identify tissue types and their accessions based on their metabolic profiles, we created a predictive Random Forest Model (Figure 3). During the training of the model, an error rate of 12.5% was estimated. Applying the final model to a testing dataset (which was not involved in the model building) resulted in the Error Matrix shown in Table 3. All five test samples were identified correctly.

Consequently, the metabolic identity of both, tissues and accessions, is sufficiently distinct to allow for a reliable classification with a Random Forest Model using LC-MS data.

### Metabolites Differentially Accumulated Among Accessions and Tissues

To better understand the variation among tissues of the different accessions we focused in the analysis on putative identified metabolites with significant differences that belong to one of the KEGG pathways (Kanehisa et al., 2014) of *A. thaliana*. With this criterion putatively identified metabolites were classified into 9 classes (Tables 1, 2), and each class was analyzed to identify whether a conserved accumulation pattern among samples and accessions was present. We also searched for changes in the presence or difference in accumulation of specific metabolites in the different tissues and/or accessions, and these are described below for each class when pertinent.

#### Class 1: Phenylpropanoids, Monolignols, and Sinapate Derivatives

Four metabolites were found in leaf and inflorescence samples that were classified as belonging to class 1 (Tables 1, 2). Interestingly, qualitative and quantitative differences were found among tissues (Tables 1, 2). Furthermore, each accession has different abundance of metabolites reflected by the fold change in the intensity of each m/z. Col-0 leaves accumulated more m/z 311.1692 and 363.0737, putatively identified as Sinapine and Sinapoyl-(S)-malate, respectively, whereas in Ws-3 leaf samples more m/z 195.0648 and 197.0803, putatively identified as Ferulic acid and 5-Hydroxyconiferyl alcohol, respectively (Table 1).

Inflorescences from Col-0 accumulated more of m/z 211.0570 and 365.1282, putatively identified as 5-Hydroxyferulic acid and Coniferin, respectively, whereas Ws-3 inflorescences

accumulated more m/z 249.0287 and 211.0931, putatively identified as Isochorismate and Sinapyl alcohol, respectively (Table 2).

#### Class 2: Prenol Lipids, Terpenoid Backbone Biosynthesis Mevalonate and MEP/DOXP Pathways

Six and nine metabolites found in leaf and inflorescence samples belong to class 2. Interestingly, qualitative differences were identified among tissues, like m/z 369.1222 that was present in leaf (Table 1), but not in inflorescence samples (Table 2). On the contrary, m/z 251.0213, 265.1432, 287.1312, and 514.2115 were detected in inflorescence, but not in leaf samples. Some of the metabolites were putatively identified as known hormones or as hormone precursors (Tables 1, 2).

In leaf samples of Col-0, the more abundant metabolite was m/z 355.1579, putatively identified as Gibberellin A20 (Table 1). The other compounds detected were accumulated more in Ws-3 leaf samples, m/z 273.1457, 333.1742, 343.2645, 355.1444, and 369.1222, which were identified as 5-(4-Hydroxy-2,2,6-trimethyl-7-oxabicyclo[4.1.0]hept-1-yl)-3-methyl-2,4-pentadienal, Gibberellin A4, 1,22-Docosane diol, Gibberellin A51, and Gibberellic acid, respectively (Table 1).

On the other hand, Col-0 inflorescence samples showed more abundance of m/z 514.2115, putatively identified as Zeatin riboside-O-glucoside, whereas Ws-3 inflorescence samples showed more accumulation of m/z 251.0213 (Mevalonate 5-phosphate), 265.1432 [(s)-(+)-Abscisic acid], 271.1319 (Abscisic acid aldehyde), 287.1312 [(s)-(+)-Abscisic acid], 331.1579 (Gibberellin A5), 333.1739 (Gibberellin A51), 343.2640 (1, 22-Docosane diol), and 354.1815 (Dihydrozeatin riboside) (Table 2).

#### Class 3: Biosynthesis of Phenylpropanoids, Biosynthesis of Flavonoids, Flavonones

Four metabolites detected in leaf and four in inflorescence samples belong to class 3. They were putatively identified as phenylpropanoids (Flavonols, Flavonones, Flavones, Anthocyanines, and Leucoanthocyanidins). Leaf and inflorescence samples showed distinctive metabolites, like m/z 311.0458 and 595.1585 that were present only in leaves (Table 1), and m/z 595.1589 and 611.1567 that were detected only in inflorescences (Table 2).

Leaves from Col-0 showed more accumulation of m/z 595.1585, putatively identified as Kaempferol-3-O-beta-D-galactoside-7-O-alpha-L-rhamnoside, while Ws-3 leaves accumulated more m/z 311.0458, 323.0307, and 329.0675, putatively identified as (+)-Dihydrokaempferol, Cyanidin 3-O-[2''-O-(2'''-O-(sinapoyl) xylosyl) glucoside] 5-O-glucoside, and Leucocyanidin, respectively (Table 1).

In contrast, Ws-3 inflorescence samples accumulated more m/z 323.0319, 329.0680, 595.1589, and 611.1567, putatively identified as Cyanidin 3-O-[2''-O-(xylosyl) glucoside] 5-O-(6'''-O-malonyl) glucoside, Leucocyanidin, Quercetin-3,7-O-a-L-dirhamnopyranoside, and Quercetin-3-O-b-glucopyranosyl-7-O-a-rhamnopyranoside, respectively.

**TABLE 1 | Relative amounts of metabolites putatively identified by ACQUITY UPLC-LCT Premier XE (Waters) in crude extracts from leaves of Ws-3 and Col-0 *Arabidopsis thaliana* accessions.**

Class	m/z	Ionization mode	Name	Chem Spider ID	Fold change (Ws-3/Col-0)	p-value
1	195.0648	[M+H] <sup>+</sup>	Ferulic acid	393368	2.4	7.85E-003
1	197.0803	[M+H] <sup>+</sup>	5-Hydroxyconiferyl alcohol	4445309	1.4	2.52E-002
1	311.1692	[M+H] <sup>+</sup>	Sinapine	80576	0.7	2.93E-005
1	363.0737	[M+Na] <sup>+</sup>	Sinapoyl-(S)-malate	4444177	0.9	7.82E-002
2	273.1457	[M+Na] <sup>+</sup>	5-(4-Hydroxy-2,2,6-trimethyl-7-oxabicyclo[4.1.0]hept-1-yl)-3-methyl-2,4-pentadienal	21172770	1.4	1.79E-003
2	333.1742	[M+H] <sup>+</sup>	Gibberellin A4	10222155	2.0	1.36E-002
2	343.2645	[M+Na] <sup>+</sup>	1,22-Docosane diol	190585	1.5	1.79E-005
2	355.1444	[M+Na] <sup>+</sup>	Gibberellin A51	391672	1.4	3.40E-002
2	355.1579	[M+Na] <sup>+</sup>	Gibberellin A20	20015789	0.6	3.72E-003
2	369.1222	[M+Na] <sup>+</sup>	Gibberellic acid	6223	1.3	6.84E-003
3	311.0458	[M+Na] <sup>+</sup>	(+)-Dihydrokaempferol	109514	2.3	1.03E-006
3	323.0307	[M+H] <sup>+</sup>	Cyanidin 3-O-[2''-O-(2'''-O-(sinapoyl) xylosyl) glucoside] 5-O-glucoside	61546	3.3	2.54E-005
3	329.0675	[M+Na] <sup>+</sup>	Leucocyanidin	64694	1.5	8.34E-003
3	595.1585	[M+H] <sup>+</sup>	Kaempferol-3-O-beta-D-galactoside-7-O-alpha-L-rhamnoside	28481780	0.7	5.40E-002
4	223.1695	[M+Na] <sup>+</sup>	Lauric acid	3756	0.4	4.25E-004
5	190.0039	[M+Na] <sup>+</sup>	Quinolinic acid	1037	0.6	2.57E-006
5	199.0751	[M+Na] <sup>+</sup>	N-hydroxy tryptamine	10391819	1.7	2.50E-004
6	223.0572	[M+Na] <sup>+</sup>	5-Methylsufinylpentyl nitrile	1363309	0.8	5.64E-002
6	235.0595	[M+H] <sup>+</sup>	2-(4'-Methylthio)butylmalic acid	24784695	2.0	5.15E-005
6	256.1438	[M+Na] <sup>+</sup>	Hexahomomethionine	21865788	0.7	1.22E-003
7	221.0315	[M+H] <sup>+</sup>	3-(1H-Imidazol-4-yl)-2-oxopropyl dihydrogen phosphate	770	1.1	1.94E-003
7	244.0489	[M+Na] <sup>+</sup>	L-Histidinol phosphate	388515	1.2	6.71E-003
7	275.1138	[M+Na] <sup>+</sup>	D-Coenzyme A	2654	1.4	7.87E-003
7	313.0989	[M+Na] <sup>+</sup>	N-Succinyl-L-2,6-diaminoheptanedioate	1160	3.1	1.36E-011
8	343.1212	[M+H] <sup>+</sup>	Sucrose	5768	0.9	2.74E-002
9	250.0505	[M+H] <sup>+</sup>	[5-Hydroxy-4-(hydroxymethyl)-6-methyl-3-pyridinyl]methyl dihydrogen phosphate	1026	1.3	3.72E-003

These metabolites were differentially accumulated among accessions and tissues.

#### Class 4: Fatty Acids, Fatty Acyls, Octadecanoids, Jasmonic Acid

One and five metabolites were found in leaf and inflorescence samples, respectively, that belong to class 4.

Col-0 leaf samples have more m/z 223.1695, putatively identified as Lauric acid (Table 1).

Col-0 inflorescence samples accumulated more m/z 209.1537 and 825.4698, putatively identified as Undecanoic acid and Arabidopside B, respectively (Table 2). Ws-3 inflorescences accumulated more m/z 233.1163, 233.1694, and 249.1849, putatively identified as (+)-Epijasmonic acid, Lauric acid, and Myristoleic acid, respectively (Table 2).

#### Class 5: Alkaloids

Only three metabolites with significant differences were found that belong to this class. Col-0 leaves accumulated more m/z 190.0039, putatively identified as Quinolinic acid, whereas Ws-3 accumulated more m/z 199.0751, putatively identified as N-hydroxy tryptamine. The m/z 363.0392 was distinct in

inflorescence samples and showed higher accumulation in Ws-3 than in the Col-0 accession.

#### Class 6: Glucosinolates Biosynthesis and Degradation

Three metabolites found in leaf and in inflorescence samples, respectively, belong to class 6.

We detected one distinct metabolite in leaf and one in inflorescence samples: m/z 235.0595, putatively identified as 2-(4'-Methylthio)butylmalic acid (Table 1), and m/z 465.0834, putatively identified as 6-Methylsulfanylhexyl glucosinolate, respectively (Table 2). Moreover, we could distinguish that Col-0 accumulated more m/z 223.0572 and 256.1438 in leaf samples (Table 1), and more m/z 223.0570 and 256.1437 in inflorescence samples (Table 2).

#### Class 7: Amino Acids and Amino Acid Metabolism

Class 7 contains metabolites involved in the biosynthesis or metabolism of amino acids. Four and eight metabolites that belong to this class were identified in leaf and inflorescence

**TABLE 2 | Relative amounts of metabolites putatively identified by ACQUITY UPLC-LCT Premier XE (Waters) in crude extracts from inflorescences of Ws-3 and Col-0 Arabidopsis thaliana accessions.**

Class	m/z	Ionization mode	Name	Chem Spider ID	Fold change (Ws-3/Col-0)	p-value
1	211.0570	[M+H] <sup>+</sup>	5-Hydroxyferulic acid	141117	0.6	1.28E-003
1	211.0931	[M+H] <sup>+</sup>	Sinapyl alcohol	4444145	1.5	1.60E-009
1	249.0287	[M+Na] <sup>+</sup>	Isochorismate	787	1.2	5.05E-007
1	365.1282	[M+Na] <sup>+</sup>	Coniferin	4444067	0.7	9.54E-004
2	251.0213	[M+Na] <sup>+</sup>	Mevalonate 5-phosphate	463	1.3	1.86E-005
2	265.1432	[M+H] <sup>+</sup>	(s)-(+)-Abscisic acid	4444418	1.3	2.69E-003
2	271.1319	[M+Na] <sup>+</sup>	Abscisic acid aldehyde	14797236	1.1	2.00E-003
2	273.1459	[M+Na] <sup>+</sup>	5-(4-Hydroxy-2,2,6-trimethyl-7-oxabicyclo[4.1.0]hept-1-yl)-3-methyl-2,4-pentadienal	21172770	1.3	2.87E-006
2	287.1312	[M+Na] <sup>+</sup>	(s)-(+)-Abscisic acid	4444418	1.7	1.79E-004
2	331.1579	[M+H] <sup>+</sup>	Gibberellin A5	4445248	1.2	7.78E-003
2	333.1739	[M+H] <sup>+</sup>	Gibberellin A51	391672	1.3	9.66E-003
2	343.2640	[M+Na] <sup>+</sup>	1,22-Docosane diol	190585	1.8	6.86E-007
2	514.2115	[M+H] <sup>+</sup>	Zeatin riboside-O-glucoside	9887971	0.9	8.68E-003
3	323.0319	[M+H] <sup>+</sup>	Cyanidin 3-O-[2''-O-(xylosyl) glucoside] 5-O-(6'''-O-malonyl) glucoside	61546	2.3	1.02E-008
3	329.0680	[M+Na] <sup>+</sup>	Leucocyanidin	64694	1.2	1.20E-004
3	595.1589	[M+H] <sup>+</sup>	Quercetin-3,7-O-a-L-dirhamnopyranoside	13095553	3.6	1.02E-008
3	611.1567	[M+H] <sup>+</sup>	Quercetin-3-O-b-glucopyranosyl-7-O-a-rhamnopyranoside	29273176	2.3	3.94E-009
4	209.1537	[M+Na] <sup>+</sup>	Undecanoic acid	7888	0.8	5.84E-005
4	223.1694	[M+Na] <sup>+</sup>	Lauric acid	3756	1.3	3.37E-004
4	233.1163	[M+Na] <sup>+</sup>	(+)-Epjasmonic acid	5584839	1.2	1.72E-002
4	249.1849	[M+Na] <sup>+</sup>	Myristoleic acid	4444564	1.4	2.82E-004
4	825.4698	[M+Na] <sup>+</sup>	Arabidopsin B	10192487	0.8	5.02E-002
5	363.0392	[M+H] <sup>+</sup>	Xanthosine 5'-monophosphate	10463791	2.0	8.22E-008
6	223.0570	[M+Na] <sup>+</sup>	5-Methylsulfanylpentyl nitrile	1363309	0.6	2.13E-004
6	256.1437	[M+Na] <sup>+</sup>	Hexahomomethionine	21865788	0.6	1.65E-002
6	465.0834	[M+H] <sup>+</sup>	6-Methylsulfanylhexyl glucosinolate	24784958	2.5	5.92E-009
7	188.0738	[M+Na] <sup>+</sup>	Phenylalanine	969	3.6	8.82E-005
7	199.0751	[M+Na] <sup>+</sup>	N-hydroxy tryptamine	10391819	1.3	1.84E-003
7	213.0917	[M+Na] <sup>+</sup>	Meso-2,6-Diaminoheptanedioate	89700	1.7	2.94E-005
7	221.0315	[M+H] <sup>+</sup>	Imidazole acetol phosphate	770	1.1	4.32E-005
7	231.0831	[M+Na] <sup>+</sup>	L-Kynurenine	141580	0.8	3.71E-002
7	244.0489	[M+Na] <sup>+</sup>	L-Histidinol phosphate	388515	1.1	5.83E-004
7	385.1300	[M+H] <sup>+</sup>	2-Amino-4-(((5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydro-2-furanyl)methyl)sulfanyl)butanoic acid	188	1.3	2.60E-004
8	328.9407	[M+Na] <sup>+</sup>	D-Ribulose 1,5-bisphosphate	19951541	0.6	1.48E-007
8	365.1075	[M+Na] <sup>+</sup>	Sucrose	5768	1.7	8.78E-005

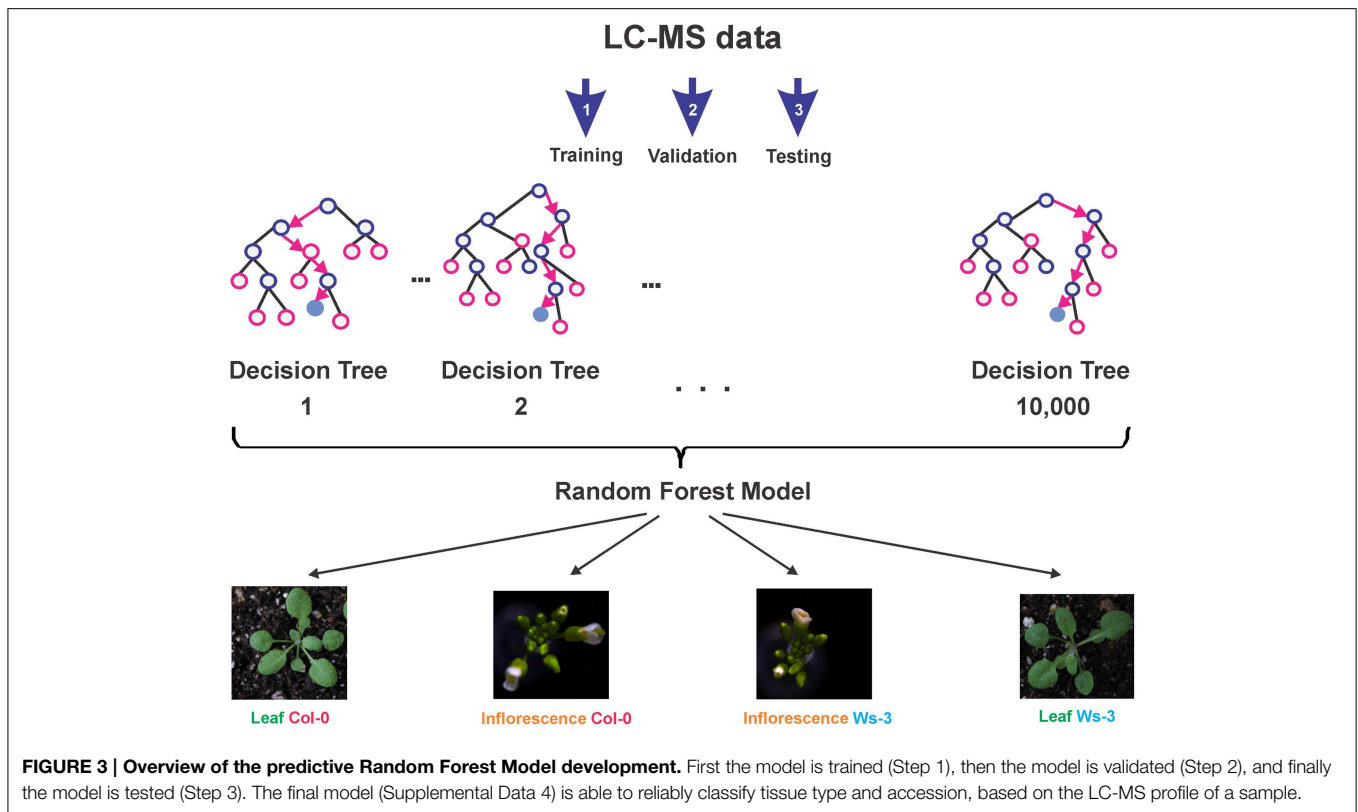
These metabolites were differentially accumulated among accessions and tissues.

samples, respectively, and only two of them were detected in both tissues.

Ws-3 leaves accumulated more of the following m/z: 221.0315, 244.0489, 275.1138, and 313.0989 (Table 1), whereas Ws-3 inflorescences accumulated more of the m/z 188.0738, 199.0751, 213.0917, 221.0315, 244.0489, 291.1242, and 385.1300 (Table 2) than the respective Col-0 tissues. On the other hand, Col-0 inflorescence samples accumulated more of m/z 231.0831 (Table 2) than Ws-3 inflorescence.

### Class 8: Carbohydrates

We putatively identified two metabolites that belong to this class and showed differential behavior among accessions. One of them was Sucrose (m/z 343.1212) that accumulated more in Col-0 than in Ws-3 leaves, but was more abundant in Ws-3 than in Col-0 inflorescences (Table 1). The second one was D-Ribulose 1,5-bisphosphate (m/z 328.9407), which was more abundant in Col-0 inflorescences samples.



## Class 9: Vitamins

We found significant variation in one metabolite that belongs to this class, *m/z* 250.0505, putatively identified as [5-Hydroxy-4-(hydroxymethyl)-6-methyl-3-pyridinyl]methyl dihydrogen phosphate, which was more abundant in Ws-3 (Table 1) than Col-0 leaves.

In summary, metabolic profiling revealed distinct metabolic phenotypes for each accession and tissue. The metabolic phenotype included metabolites from at least 9 different classes. Figures 4, 5, as an example, represent some of the metabolic differences observed in leaf and inflorescence samples. Besides the potential use of these metabolites for identification of accession and tissues, some of them like sucrose, gibberellins A20, D-Ribulose 1,5-biphosphate, are interesting for further studies that could help to understand the morphological differences as well as the growth potential of the accessions.

## Discussion

### Distinct Metabolic Phenotypes Were Detected for Different Accessions and Tissues

Different metabolomic approaches have been undertaken seeking to establish a link between genotypes and phenotypes, e.g., some studies have described the variation of individual and specific classes of metabolites and the genetics involved in their control (Kliebenstein et al., 2001b; Wentzell et al., 2007; Chan et al., 2009), while others have described the natural variation of the

**TABLE 3 | An Error Matrix demonstrates the performance of the Random Forest Model for the correct classification of tissue type and accession of Arabidopsis thaliana.**

Actual	Predicted			
	Col-0_Inflor	Col-0_Leaf	Ws-3_Inflor	Ws-3_Leaf
Col-0_Inflor	<b>1</b>	0	0	0
Col-0_Leaf	0	<b>2</b>	0	0
Ws-3_Inflor	0	0	<b>1</b>	0
Ws-3_Leaf	0	0	0	<b>1</b>

*Inflor*: Inflorescence.

The bold values indicate that the actual identities of the samples are equal to the predictions.

metabolome of Arabidopsis through untargeted metabolomics analysis (Keurentjes et al., 2006; Rowe et al., 2008; Joseph et al., 2013, 2014).

Those studies uncovered qualitative and quantitative variation in metabolite accumulation between Arabidopsis accessions. The combination of metabolic profiling and genetics allowed the identification of QTLs associated for about 75% of the mass signals (Keurentjes et al., 2006). Other studies showed that metabolite QTLs were organized in 6 clusters with phenotypic effects (Rowe et al., 2008; Fu et al., 2009). Some of these clusters were associated with central metabolism (Rowe et al., 2008) and others to previously identified secondary metabolite loci (Kliebenstein et al., 2001a,b; Hansen et al., 2007; Wentzell et al.,



2007). Furthermore, there is evidence that different factors like environment, tissue type, and plant age affect the outcomes of the genetic network controlling metabolism in *Arabidopsis* (Wentzell et al., 2008; Wentzell and Kliebenstein, 2008).

In our study we observed natural metabolic variation among two different tissues of the accessions Col-0 and Ws-3, which are commonly used in research. The mass profiles of each tissue of the different accessions presented quantitative and qualitative variation, allowing us to distinguish among these accessions and tissues in terms of their metabolic profiles.

We detected more than 14,000 and 17,000 peaks from inflorescences and leaves, respectively, in the two accessions. 222 high quality features presented significant differences ( $p \leq 0.01$ ) in leaf samples and 418 high quality features did in inflorescence samples ( $p \leq 0.01$ ). From these metabolites that presented significant differences we could putatively identify 26 and 36 metabolites in leaf and in inflorescence samples (Tables 1, 2), with 17 of those metabolites present in both tissues in the two accessions (Supplementary Table S1). Although many signals remain unidentified, we created a Random Forest Model, which permits the classification of both tissue and accession based on their metabolic fingerprint. The model is predictive and

may be employed for the correct identification of otherwise indistinguishable plants.

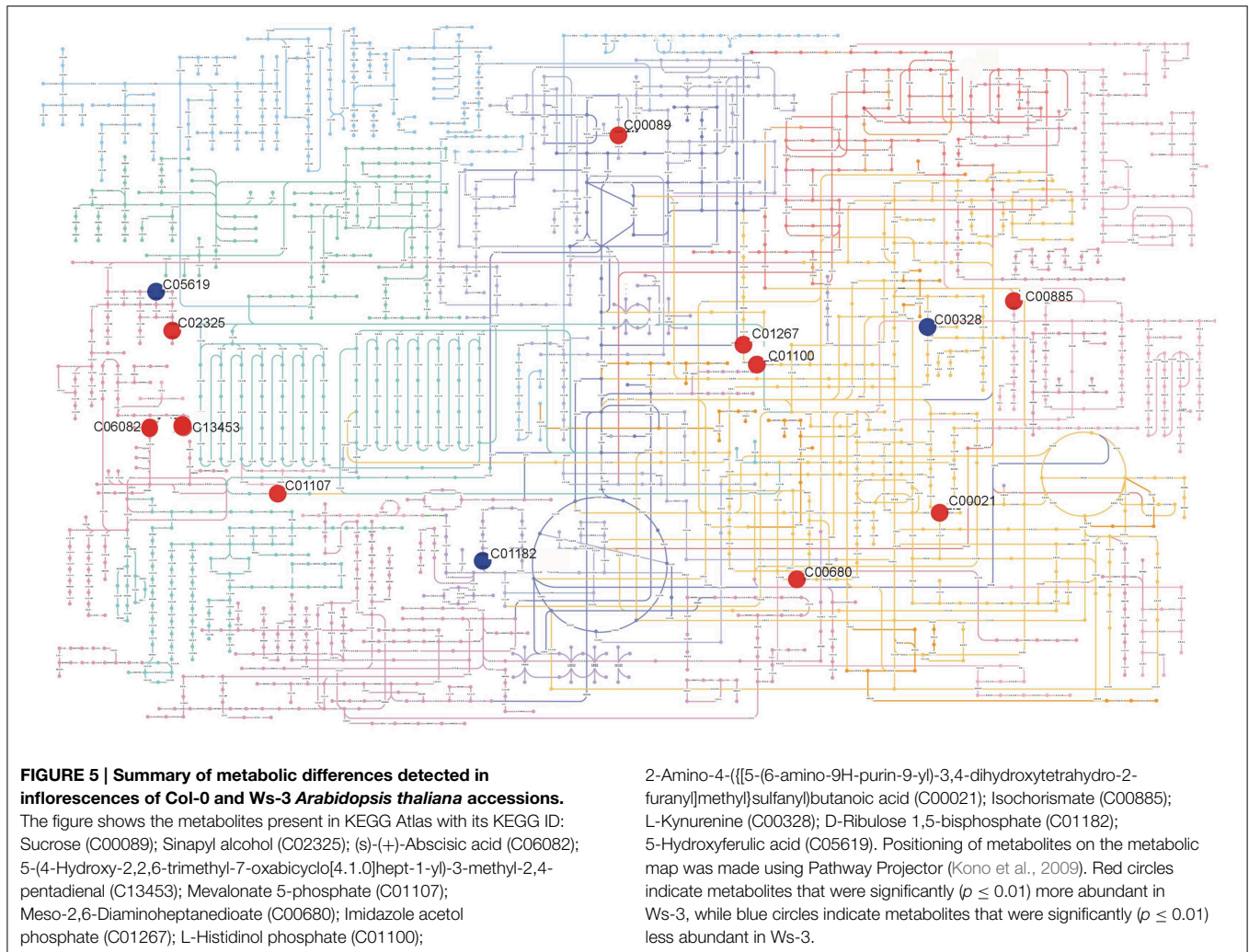
### Nine Metabolite Classes Were Differentially Accumulated in Different Accessions and Tissues

In this study we found quantitative variation in nine metabolite classes, indicating different metabolite compositions in each accession and tissue.

Several studies using *A. thaliana* natural accessions have shown that differential gene expression exists among the accessions. The most differentially expressed genes concerned to the response to the biotic environment, including pathogen defense and the production of glucosinolates (West et al., 2006, 2007; Kliebenstein et al., 2006a,b; van Leeuwen et al., 2007; Gan et al., 2011). In agreement with these observations, our study showed quantitative and qualitative variation in metabolites related to pathogen defense.

The glucosinolates are a group of naturally occurring metabolites in the Brassicales order involved in plant defense (Wittstock and Burow, 2010). These thioglucosides are derived from a variety of protein amino acids (Met, Leu, iso-Leu, Val, Trp, and Phe) (Kliebenstein et al., 2001b; Keurentjes et al., 2006).





Particularly, in our extraction conditions, we observed one distinct metabolite for leaves ( $m/z$  235.0595; **Table 1**) and one for inflorescences ( $m/z$  465.0834; **Table 2**), and a different accumulation pattern for certain metabolites in different accessions:  $m/z$  223.0572 and 256.1438 (**Table 1**) were more abundant in Col-0 leaf samples, and  $m/z$  223.0570 and 256.1437 (**Table 2**) in Col-0 inflorescences. These metabolites were putatively identified to participate in glucosinolate biosynthesis and degradation.

Other studies reported glucosinolate variation in leaves and seeds of *Arabidopsis* accessions (Kliebenstein et al., 2001b; Matsuda et al., 2010), and they have been used to discriminate among some *A. petraea* populations (Davey et al., 2008). They also have been shown to be subjected to genetic variation in *Arabidopsis* (Kliebenstein et al., 2001b).

Other metabolites, classified in our study as belonging to class 4, were differentially accumulated among tissues and accessions. Ws-3 inflorescences accumulated more (+)-Epijasmonic acid ( $m/z$  233.1163; **Table 2**) than Col-0, while Col-0 inflorescences more Arabidopside B ( $m/z$  825.4698; **Table 2**). Jasmonic acid and Methyl jasmonate play an essential

role in plant defense responses, pollen development and leaf growth control by repressing cell proliferation (Świątek et al., 2004; Zhang and Turner, 2008; Chen et al., 2013; Noir et al., 2013). Arabidopside B seems to have an inhibitory effect on root growth and a possible role as a reservoir for slow release of free OPDA, a Jasmonate precursor (Kourtchenko et al., 2007).

Also, it was interesting to note the differential accumulation of Lauric acid ( $m/z$  223.1696) in the leaves of the two accessions, being Col-0 the one that accumulated the most in leaves and Ws-3 in inflorescences. It has been demonstrated that Lauric acid can be elongated and desaturated into Linolenic acid that then is incorporated into Jasmonic acid and Methyljasmonate (Afitlhile et al., 2004).

Phenylpropanoid pathway metabolites are also known for their protective roles (Buer et al., 2010; Fraser and Chapple, 2011). We observed a distinct pattern of accumulation of metabolites belonging to the flavonoid branch of this pathway among tissues and accessions, inflorescences being the samples that presented the most diversity in these compounds. Many of these compounds have also been considered as chemical

messengers, physiological regulators and cell cycle inhibitors (Buer et al., 2010; Falcone Ferreyra et al., 2012). Furthermore, a distinct pattern of accumulation of lignin precursors was also identified here as well as differences in the content of Sinapate esters.

The rate of plant growth depends on a combination of photosynthetic carbon (C) assimilation rate and developmental programs that determine how rapidly metabolites are used for growth, although the molecular and genetic basis are not well-understood (Sulpice et al., 2009). It has been reported that natural variation in the level of central metabolites exists (Loudet et al., 2003; Calenge et al., 2006) and that there are positive and negative correlations between these metabolites and biomass (Meyer et al., 2007; Sulpice et al., 2010). Metabolites that are negatively correlated with biomass were sucrose, glucose- and fructose-6-phosphate, which link carbon flow from photosynthesis and starch and sucrose metabolism with cell wall formation, the TCA cycle members citric acid, succinate or malic acid, as well as the amino acids glutamine and phenylalanine (Meyer et al., 2007). In our study, we observed differences in sucrose content in leaf and inflorescence samples between accessions, and of D-ribulose 1,5 biphosphate in inflorescence samples as well as differences in the content of phenylalanine and phenylalanine derived compounds that contribute to cell wall formation (ferulic acid, sinapine). Houshyani et al. (2012) observed significant differences among Arabidopsis accessions for some primary metabolites, e.g., fructose, 1-methyl- $\alpha$ -D-glucopyranoside, glucopyranose, sucrose, and L-glutamic acid. Metabolite QTLs were also associated with central metabolism, suggesting that differences in central carbon metabolism can exist among accessions (Rowe et al., 2008; Houshyani et al., 2012).

Plant hormones affect gene expression and transcription levels promoting cellular division, growth and differentiation, and directing developmental programs that determine how rapidly metabolites are used for growth (Alabadi et al., 2009; Sulpice et al., 2010).

Paparelli et al. (2013) showed that plant size is determined by a mechanism in which carbohydrates produced by photosynthesis modulate the synthesis of gibberellins (Paparelli et al., 2013). In our study, we observed a different pattern of GAs and sucrose accumulation among accessions (Tables 1, 2 and Figures 4, 5), which could be further investigated to establish a correlation with the growth phenotypes.

Moreover, differences between accessions in the content of abscisic acid or intermediates in abscisic acid biosynthesis were found. Abscisic acid is well-known for its effect on seed germination, flowering, and during plant response to environmental stress and plant pathogens. Col-0 and Ws-3 present differences in flowering time that could eventually be explained by the differences here found in the content of GAs

and ABA, hormones that concur to regulate this process. Recent studies showed that ABA potentially delays flowering under unstressed conditions, but promotes it when plants are stressed (Finkelstein, 2013).

Further work must be done to investigate whether there is a correlation between the hormonal differences and the metabolic signatures found in each accession that could explain the morphological differences among accessions.

## Concluding Remarks

In this work we found a distinct metabolic phenotype of each tissue and Arabidopsis accession studied. We found quantitative variation in nine metabolite classes, resulting in different compositions of metabolites in each accession and tissue. Moreover, a predictive Random Forest Model was made that is able to reliably classify tissue type and accession of samples based on LC-MS profiles.

The metabolite signature of accessions found in this work could set a basis for future studies to understand how these profiles correlate with their respective phenotype. For example, by exploring its correlation with interesting developmental processes, like cell division, cell expansion, flowering time, and biomass production. Moreover, knowledge of metabolite natural diversity could help to direct plant breeding approaches.

## Author Contributions

MS performed the sample preparations and AC performed chromatography and mass spectrometry experiments. MS, NM, RW, and SF conceived the project, designed the experiments, and analyzed the data. RW performed data processing, Random Forest Model development, and statistical analysis. MS, NM, RW, and SF drafted the manuscript. All authors read and approved the final manuscript.

## Acknowledgments

We would like to thank the two reviewers for their constructive comments. MS was supported by the Mexican National Council of Science and Technology (CONACyT) fellowship (229496). SF acknowledges Langebio intramural funds and the CONACyT grant CB-2012-01/177739, NMM CONACyT grant CB-2011-01/165986, and RW CONACyT grant CB-2010-01/151596.

## Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2015.00365/abstract>

## References

- Afendi, F. M., Okada, T., Yamazaki, M., Hirai-Morita, A., Nakamura, Y., Nakamura, K., et al. (2012). KnapSack family databases: integrated metabolite-plant species databases for multifaceted plant research. *Plant Cell Physiol.* 53:e1. doi: 10.1093/pcp/pcr165
- Afithile, M. M., Fukushige, H., and Hildebrand, D. (2004). Labeling of major plant lipids and jasmonic acid using [1-<sup>14</sup>C] lauric acid. *Phytochemistry* 65, 2679–2684. doi: 10.1016/j.phytochem.2004.08.028
- AGI. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815. doi: 10.1038/35048692
- Alabadi, D., Blazquez, M. A., Carbonell, J., Ferrandiz, C., and Perez-Amador, M. A. (2009). Instructive roles for hormones in plant development. *Int. J. Dev. Biol.* 53, 1597–1608. doi: 10.1387/ijdb.072423da
- Alonso-Blanco, C., and Koornneef, M. (2000). Naturally occurring variation in *Arabidopsis*: an underexploited resource for plant genetics. *Trends Plant Sci.* 5, 22–29. doi: 10.1016/S1360-1385(99)01510-1
- Alonso-Blanco, C., Blankestijn-de Vries, H., Hanhart, C. J., and Koornneef, M. (1999). Natural allelic variation at seed size loci in relation to other life history traits of *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 96, 4710–4717. doi: 10.1073/pnas.96.8.4710
- Barah, P., Jayavelu, N., Rasmussen, S., Nielsen, H., Mundy, J., and Bones, A. (2013). Genome-scale cold stress response regulatory networks in ten *Arabidopsis thaliana* ecotypes. *BMC Genomics* 14:722. doi: 10.1186/1471-2164-14-722
- Beemster, G. T., De Vusser, K., De Tavernier, E., De Bock, K., and Inze, D. (2002). Variation in growth rate between *Arabidopsis* ecotypes is correlated with cell division and A-type cyclin-dependent kinase activity. *Plant Physiol.* 129, 854–864. doi: 10.1104/pp.002923
- Buer, C. S., Imin, N., and Djordjevic, M. A. (2010). Flavonoids: new roles for old molecules. *J. Integr. Plant Biol.* 52, 98–111. doi: 10.1111/j.1744-7909.2010.00905.x
- Calenge, F., Saliba-Colombani, V. R., Mahieu, S. P., Loudet, O., Daniel-Vedele, F. O., and Krapp, A. (2006). Natural variation for carbohydrate content in *Arabidopsis*. Interaction with complex traits dissected by quantitative genetics. *Plant Physiol.* 141, 1630–1643. doi: 10.1104/pp.106.082396
- Chambers, M. C., MacLean, B., Burke, R., Amodei, D., Ruderman, D. L., Neumann, S., et al. (2012). A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotech.* 30, 918–920. doi: 10.1038/nbt.2377
- Chan, E. K. F., Rowe, H. C., and Kliebenstein, D. J. (2009). Understanding the evolution of defense metabolites in *Arabidopsis thaliana* using genome-wide association mapping. *Genetics* 185, 991–1007. doi: 10.1534/genetics.109.108522
- Chan, E. K. F., 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
- Chen, J., Sonobe, K., Ogawa, N., Masuda, S., Nagatani, A., Kobayashi, Y., et al. (2013). Inhibition of *Arabidopsis* hypocotyl elongation by jasmonates is enhanced under red light in phytochrome B dependent manner. *J. Plant Res.* 126, 161–168. doi: 10.1007/s10265-012-0509-3
- Chevalier, F., Martin, O., Rofidal, V., Devauchelle, A. D., Barteau, S., Sommerer, N., et al. (2004). Proteomic investigation of natural variation between *Arabidopsis* ecotypes. *Proteomics* 4, 1372–1381. doi: 10.1002/pmic.200300750
- Chevalier, F., Pata, M., Nacry, P., Doumas, P., and Rossignol, M. (2003). Effects of phosphate availability on the root system architecture: large-scale analysis of the natural variation between *Arabidopsis* accessions. *Plant Cell Environ.* 26, 1839–1850. doi: 10.1046/j.1365-3040.2003.01100.x
- Davey, M. P., Burrell, M. M., Woodward, F. I., and Quick, W. P. (2008). Population-specific metabolic phenotypes of *Arabidopsis lyrata* ssp. *petraea*. *New Phytol.* 177, 380–388. doi: 10.1111/j.1469-8137.2007.02282.x
- Enot, D. P., and Draper, J. (2007). Statistical measures for validating plant genotype similarity assessments following multivariate analysis of metabolome fingerprint data. *Metabolomics* 3, 349–355. doi: 10.1007/s11306-007-0066-2
- Ernest, B., Gooding, J. R., Campagna, S. R., Saxton, A. M., and Voy, B. H. (2012). MetabR: an R script for linear model analysis of quantitative metabolomic data. *BMC Res. Notes* 5:596. doi: 10.1186/1756-0500-5-596
- Falcone Ferreyra, M. L., Rius, S., and Casati, P. (2012). Flavonoids: biosynthesis, biological functions and biotechnological applications. *Front. Plant Sci.* 3:222. doi: 10.3389/fpls.2012.00222
- Fernie, A. R., and Schauer, N. (2009). Metabolomics-assisted breeding: a viable option for crop improvement? *Trends Genet.* 25, 39–48. doi: 10.1016/j.tig.2008.10.010
- Fiehn, O., Kopka, J., Dormann, P., Altmann, T., Trethewey, R. N., and Willmitzer, L. (2000). Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* 18, 1157–1161. doi: 10.1038/81137
- Finkelstein, R. (2013). Abscisic acid synthesis and response. *Arabidopsis Book* 11:e0166. doi: 10.1199/tab.0166
- Fraser, C. M., and Chapple, C. (2011). The phenylpropanoid pathway in *Arabidopsis*. *Arabidopsis Book* 9:e0152. doi: 10.1199/tab.0152
- Fu, J., Keurentjes, J. J. B., Bouwmeester, H., America, T., Verstappen, F. W. A., Ward, J. L., et al. (2009). System-wide molecular evidence for phenotypic buffering in *Arabidopsis*. *Nat. Genet.* 41, 166–167. doi: 10.1038/ng.308
- Gan, X., Stegle, O., Behr, J., Steffen, J. G., Drewe, P., Hildebrand, K. L., et al. (2011). Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. *Nature* 477, 419–423. doi: 10.1038/nature10414
- García-Flores, M., Juárez-Colunga, S., Garca-Casarrubias, A., Trachsel, S., Winkler, R., and Tiessen, A. (2015). Metabolic profiling of plant extracts using direct-injection electrospray ionization mass spectrometry allows for high-throughput phenotypic characterization according to genetic and environmental effects. *J. Agric. Food Chem.* 63, 1042–1052. doi: 10.1021/jf504853w
- García-Flores, M., Juárez-Colunga, S., Montero-Vargas, J. M., López-Arciniega, J. A. I., Chagolla, A., Tiessen, A., et al. (2012). Evaluating the physiological state of maize (*Zea mays* L.) plants by direct-injection electrospray mass spectrometry (DIESI-MS). *Mol. Biosyst.* 8, 1658–1660. doi: 10.1039/c2mb25056j
- Gaut, B. (2012). *Arabidopsis thaliana* as a model for the genetics of local adaptation. *Nat. Genet.* 44, 115–116. doi: 10.1038/ng.1079
- Hansen, B. G., Kliebenstein, D. J., and Halkier, B. A. (2007). Identification of a flavin-monooxygenase as the S-oxygenating enzyme in aliphatic glucosinolate biosynthesis in *Arabidopsis*. *Plant J.* 50, 902–910. doi: 10.1111/j.1365-313X.2007.03101.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
- Houshyani, B., Kabouw, P., Muth, D., de Vos, R. H., Bino, R., and Bouwmeester, H. (2012). Characterization of the natural variation in *Arabidopsis thaliana* metabolome by the analysis of metabolic distance. *Metabolomics* 8, 131–145. doi: 10.1007/s11306-011-0375-3
- Joseph, B., Atwell, S., Corwin, J., Li, B., and Kliebenstein, D. J. (2014). Meta-analysis of Metabolome QTLs in *Arabidopsis*: trying to estimate the network size controlling genetic variation of the metabolome. *Front. Plant Sci.* 5:461. doi: 10.3389/fpls.2014.00461
- Joseph, B., Corwin, J. A., Li, B., Atwell, S., and Kliebenstein, D. J. (2013). Cytoplasmic genetic variation and extensive cytonuclear interactions influence natural variation in the metabolome. *Elife* 2:e00776. doi: 10.7554/eLife.00776
- Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* 42, D199–D205. doi: 10.1093/nar/gkt1076
- Keurentjes, J. J., Fu, J., de Vos, C. H., Lommen, A., Hall, R. D., Bino, R. J., et al. (2006). The genetics of plant metabolism. *Nat. Genet.* 38, 842–849. doi: 10.1038/ng1815
- Keurentjes, J. J., Sulpice, R., Gibon, Y., Steinhauser, M. C., Fu, J., Koornneef, M., et al. (2008). Integrative analyses of genetic variation in enzyme activities of primary carbohydrate metabolism reveal distinct modes of regulation in *Arabidopsis thaliana*. *Genome Biol.* 9:R129. doi: 10.1186/gb-2008-9-8-r129
- Kliebenstein, D. J., Gershenzon, J., and Mitchell-Olds, T. (2001a). Comparative quantitative trait loci mapping of aliphatic, indolic and benzylic glucosinolate production in *Arabidopsis thaliana* leaves and seeds. *Genetics* 159, 359–370.
- Kliebenstein, D. J., Kroymann, J., Brown, P., Figuth, A., Pedersen, D., Gershenzon, J., et al. (2001b). Genetic control of natural variation in *Arabidopsis* glucosinolate accumulation. *Plant Physiol.* 126, 811–825. doi: 10.1104/pp.126.2.811
- Kliebenstein, D. J., West, M. A., van Leeuwen, H., Kim, K., Doerge, R. W., Michelsmore, R. W., et al. (2006a). Genomic survey of gene

- expression diversity in *Arabidopsis thaliana*. *Genetics* 172, 1179–1189. doi: 10.1534/genetics.105.049353
- Kliebenstein, D. J., West, M. A., van Leeuwen, H., Loudet, O., Doerge, R. W., and St. Clair, D. A. (2006b). Identification of QTLs controlling gene expression networks defined a priori. *BMC Bioinformatics* 7:308. doi: 10.1186/1471-2105-7-308
- Kono, N., Arakawa, K., Ogawa, R., Kido, N., Oshita, K., Ikegami, K., et al. (2009). Pathway projector: web-based zoomable pathway browser using KEGG Atlas and Google Maps. *A. P. I. PLoS ONE* 4:e7710. doi: 10.1371/journal.pone.0007710
- 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
- Kourtchenko, O., Andersson, M. X., Hamberg, M., Brunnstrom, A., Gobel, C., McPhail, K. L., et al. (2007). Oxo-phytydienoic acid-containing galactolipids in *Arabidopsis*: jasmonate signaling dependence. *Plant Physiol.* 145, 1658–1669. doi: 10.1104/pp.107.104752
- Krueger, S., Giavalisco, P., Krall, L., Steinhauser, M.-C., Büssis, D., Usadel, B., et al. (2011). A topological map of the compartmentalized *Arabidopsis thaliana* leaf metabolome. *PLoS ONE* 6:e17806. doi: 10.1371/journal.pone.0017806
- Laibach, F. (1951). Über Sommer und Winterannuelle Rasse von *Arabidopsis thaliana* (L.) Heynh. Ein Beitrag zur Atiologie der Blütenbildung. *Beitr. Biol. Pflanzten* 28, 173–210.
- Landesfeind, M., Kaever, A., Feussner, K., Thurow, C., Gatz, C., Feussner, I., et al. (2014). Integrative study of *Arabidopsis thaliana* metabolomic and transcriptomic data with the interactive MarVis-Graph software. *PeerJ* 2:e239. doi: 10.7717/peerj.239
- Loudet, O., Chaillou, S., Merigout, P., Talbotec, J. L., and Daniel-Vedele, F. O. (2003). Quantitative trait loci analysis of nitrogen use efficiency in *Arabidopsis*. *Plant Physiol.* 131, 345–358. doi: 10.1104/pp.102.010785
- Maloo, J. N., Borevitz, J. O., Dabi, T., Lutes, J., Nehring, R. B., Redfern, J. L., et al. (2001). Natural variation in light sensitivity of *Arabidopsis*. *Nat. Genet.* 29, 441–446. doi: 10.1038/ng777
- Massonnet, C., Vile, D., Fabre, J., Hannah, M. A., Caldana, C., Lise, J., et al. (2010). Probing the reproducibility of leaf growth and molecular phenotypes: a comparison of three *Arabidopsis* accessions cultivated in ten laboratories. *Plant Physiol.* 152, 2142–2157. doi: 10.1104/pp.109.148338
- Matsuda, F., Hirai, M. Y., Sasaki, E., Akiyama, K., Yonekura-Sakakibara, K., Provart, N. J., et al. (2010). AtMetExpress development: a phytochemical atlas of *Arabidopsis* development. *Plant Physiol.* 152, 566–578. doi: 10.1104/pp.109.148031
- Matsuda, F., Nakabayashi, R., Sawada, Y., Suzuki, M., Hirai, M. Y., Kanaya, S., et al. (2011). Mass spectra-based framework for automated structural elucidation of metabolome data to explore phytochemical diversity. *Front. Plant Sci.* 2:40. doi: 10.3389/fpls.2011.00040
- Matsuda, F., Yonekura-Sakakibara, K., Niida, R., Kuromori, T., Shinozaki, K., and Saito, K. (2009). MS/MS spectral tag-based annotation of non-targeted profile of plant secondary metabolites. *Plant J.* 57, 555–577. doi: 10.1111/j.1365-3113X.2008.03705.x
- Meyer, R. C., Steinfath, M., Lise, 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
- Mintz-Oron, S., Meir, S., Malitsky, S., Rupp, E., Aharoni, A., and Shlomi, T. (2012). Reconstruction of *Arabidopsis* metabolic network models accounting for subcellular compartmentalization and tissue-specificity. *Proc. Natl. Acad. Sci. U.S.A.* 109, 339–344. doi: 10.1073/pnas.1100358109
- Montero-Vargas, J. M., González-González, L. H., Gálvez-Ponce, E., Ramírez-Chávez, E., Molina-Torres, J., Chagolla, A., et al. (2013). Metabolic phenotyping for the classification of coffee trees and the exploration of selection markers. *Mol. Biosyst.* 9, 693–699. doi: 10.1039/c3mb25509c
- Noir, S., Bömer, M., Takahashi, N., Ishida, T., Tsui, T.-L., Balbi, V., et al. (2013). Jasmonate controls leaf growth by repressing cell proliferation and the onset of endoreduplication while maintaining a potential stand-by mode. *Plant Physiol.* 161, 1930–1951. doi: 10.1104/pp.113.214908
- Paparelli, E., Parlanti, S., Gonzali, S., Novi, G., Mariotti, L., Ceccarelli, N., et al. (2013). Nighttime sugar starvation orchestrates gibberellin biosynthesis and plant growth in *Arabidopsis*. *Plant Cell* 25, 3760–3769. doi: 10.1105/tpc.113.115519
- Rédei, G. P. (1992). “A heuristic glance at the past of *Arabidopsis* genetics,” in *Methods in Arabidopsis Research*, eds C. Koncz, N. H. Chua, and J. Schell (Singapore: World Scientific), 1–15.
- Rowe, H. C., Hansen, B. G., Halkier, B. A., and Kliebenstein, D. J. (2008). Biochemical networks and epistasis shape the *Arabidopsis thaliana* metabolome. *Plant Cell* 20, 1199–1216. doi: 10.1105/tpc.108.058131
- 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
- Schneeberger, K., Ossowski, S., Ott, F., Klein, J. D., Wang, X., Lanz, C., et al. (2011). Reference-guided assembly of four diverse *Arabidopsis thaliana* genomes. *Proc. Natl. Acad. Sci. U.S.A.* 108, 10249–10254. doi: 10.1073/pnas.1107739108
- Sotelo-Silveira, M., Cucinotta, M., Chauvin, A. L., Chavez Montes, R. A., Colombo, L., Marsch-Martinez, N., et al. (2013). Cytochrome P450 CYP78A9 is involved in *Arabidopsis* reproductive development. *Plant Physiol.* 162, 779–799. doi: 10.1104/pp.113.218214
- Stein, R. J., and Waters, B. M. (2012). Use of natural variation reveals core genes in the transcriptome of iron-deficient *Arabidopsis thaliana* roots. *J. Exp. Bot.* 63, 1039–1055. doi: 10.1093/jxb/err343
- Sturm, M., Bertsch, A., Gropl, C., Hildebrandt, A., Hussong, R., Lange, E., et al. (2008). OpenMS – an open-source software framework for mass spectrometry. *BMC Bioinformatics* 9:163. doi: 10.1186/1471-2105-9-163
- Sulpice, R., Pyl, E.-T., Ishihara, H., Trenkamp, S., Steinfath, M., Witucka-Wall, H., et al. (2009). Starch as a major integrator in the regulation of plant growth. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10348–10353. doi: 10.1073/pnas.0903478106
- Sulpice, R., Trenkamp, S., Steinfath, M., Usadel, B., Gibon, Y., Witucka-Wall, H., et al. (2010). Network analysis of enzyme activities and metabolite levels and their relationship to biomass in a large panel of *Arabidopsis* accessions. *Plant Cell* 22, 2872–2893. doi: 10.1105/tpc.110.076653
- Świątek, A., Dongen, W. V., and Esmans, E. L., Onckelen, H. V. (2004). Metabolic fate of jasmonates in tobacco bright yellow-2 cells. *Plant Physiol.* 135, 161–172. doi: 10.1104/pp.104.040501
- Taylor, J., King, R. D., Altmann, T., and Fiehn, O. (2002). Application of metabolomics to plant genotype discrimination using statistics and machine learning. *Bioinformatics* 18, S241–S248. doi: 10.1093/bioinformatics/18.suppl\_2.S241
- van Leeuwen, H., Kliebenstein, D. J., West, M. A. L., Kim, K., van Poecke, R., Katagiri, F., et al. (2007). Natural variation among *Arabidopsis thaliana* accessions for transcriptome response to exogenous salicylic acid. *Plant Cell* 19, 2099–2110. doi: 10.1105/tpc.107.050641
- Wang, Y., Yang, L., Zheng, Z., Grumet, R., Loeschner, W., Zhu, J.-K., et al. (2013). Transcriptomic and physiological variations of three *Arabidopsis* ecotypes in response to salt stress. *PLoS ONE* 8:e69036. doi: 10.1371/journal.pone.0069036
- Weigel, D. (2012). Natural variation in *Arabidopsis*: from molecular genetics to ecological genomics. *Plant Physiol.* 158, 2–22. doi: 10.1104/pp.111.189845
- Wentzell, A. M., and Kliebenstein, D. J. (2008). Genotype, age, tissue, and environment regulate the structural outcome of glucosinolate activation. *Plant Physiol.* 147, 415–428. doi: 10.1104/pp.107.115279
- Wentzell, A. M., Boeye, I., Zhang, Z., and Kliebenstein, D. J. (2008). Genetic networks controlling structural outcome of glucosinolate activation across development. *PLoS Genet.* 4:e1000234. doi: 10.1371/journal.pgen.1000234
- Wentzell, A. M., Rowe, H. C., Hansen, B. G., Ticconi, C., Halkier, B. A., and Kliebenstein, D. J. (2007). Linking metabolic QTLs with network and cis-eQTLs controlling biosynthetic pathways. *PLoS Genet.* 3:e162. doi: 10.1371/journal.pgen.0030162
- West, M. A. L., Kim, K., Kliebenstein, D. J., van Leeuwen, H., Michelmore, R. W., Doerge, R. W., et al. (2007). Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in *Arabidopsis*. *Genetics* 175, 1441–1450. doi: 10.1534/genetics.106.064972
- West, M. A. L., van Leeuwen, H., Kozik, A., Kliebenstein, D. J., Doerge, R. W., St. Clair, D. A., et al. (2006). High-density haplotyping with microarray-based expression and single feature polymorphism markers in *Arabidopsis*. *Genome Res.* 16, 787–795. doi: 10.1101/gr.5011206

- Williams, G. (1987). Some experiments in decision tree induction. *Aust. Comput. J.* 19, 84–91.
- Williams, G. (1988, July). Artificial intelligence developments and applications 1987: conference papers. *Elsevier Science Ltd*.
- Williams, G. (2009). Rattle: a data mining GUI for R. *R J.* 1, 45–55.
- Williams, G. (2011). *Data Mining with Rattle and R: The Art of Excavating Data for Knowledge Discovery (Use R!)*. New York, NY: Springer-Verlag.
- Winkler, R. (2014). MASSyPup—an ‘out of the box’ solution for the analysis of mass spectrometry data. *J. Mass Spectrom.* 49, 37–42. doi: 10.1002/jms.3314
- Winkler, R. (2015). SpiderMass: semantic database creation and tripartite metabolite identification strategy. *J. Mass Spectrom.* 50, 538–541. doi: 10.1002/jms.3559
- Wittstock, U., and Burow, M. (2010). Glucosinolate breakdown in Arabidopsis: mechanism, regulation and biological significance. *Arabidopsis Book* 8:e0134. doi: 10.1199/tab.0134
- Zhang, Y., and Turner, J. G. (2008). Wound-induced endogenous jasmonates stunt plant growth by inhibiting mitosis. *PLoS ONE* 3:e3699. doi: 10.1371/journal.pone.0003699

**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 © 2015 Sotelo-Silveira, Chauvin, Marsch-Martínez, Winkler and de Folter. 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.