SCIENTIFIC DATA

OPEN

Data Descriptor: The natural variance of the Arabidopsis floral secondary metabolites

Takayuki Tohge[1,2], Monica Borghi[1] & Alisdair R. Fernie[1]

Received: 12 October 2017
Accepted: 7 February 2018
Published: 3 April 2018

Application of mass spectrometry-based metabolomics enables the detection of genotype-related natural variance in metabolism. Differences in secondary metabolite composition of flowers of 64 *Arabidopsis thaliana* (Arabidopsis) natural accessions, representing a considerable portion of the natural variation in this species are presented. The raw metabolomic data of the accessions and reference extracts derived from flavonoid knockout mutants have been deposited in the MetaboLights database. Additionally, summary tables of floral secondary metabolite data are presented in this article to enable efficient re-use of the dataset either in metabolomics cross-study comparisons or correlation-based integrative analysis of other metabolomic and phenotypic features such as transcripts, proteins and growth and flowering related phenotypes.

Design Type(s)	parallel group design • individual genetic characteristics comparison design
Measurement Type(s)	metabolite profiling • transcription profiling by array assay
Technology Type(s)	mass spectrometry assay • microarray
Factor Type(s)	geographic location • selectively maintained organism
Sample Characteristic(s)	Arabidopsis thaliana • flower • Austria • Belgium • Czech republic • Estonia • Finland • French Republic • Germany • India • Kingdom of Denmark • Kingdom of Norway • Kingdom of Spain • Kingdom of the Netherlands • Libya • Lithuania • Morocco • Poland • Portuguese Republic • Republic of Ireland • Russia • Senegal • Sweden • Switzerland • Tajikistan • Ukraine • United Kingdom • United States of America

[1]Max-Planck-Institute of Molecular Plant Physiology, 14476 Potsdam-Golm, Germany. [2]Graduate School of Biological Sciences, Nara Institute of Science and Technology, Ikoma, Nara, 630-0192, Japan. Correspondence and requests for materials should be addressed to T.T. (email: tohge@bs.naist.jp) or A.R.F. (email: fernie@mpimp-golm.mpg.de).

Background and Summary

Plant secondary metabolites (so-called specialized metabolites) that have high natural diversity in their chemical structures and abundances can be identified through metabolic screening of populations even in the comparisons between ecotypes and cultivars belonging to the same species[1–3]. This may represent relatively recent adaptations or more phylogenetical restrictions in the evolution of such metabolisms[3–5]. With metabolomic screening of such populations, metabolic polymorphism in aliphatic glucosinolates[6], flavonol-glycosides[7] and phenylacylated-flavonols[3] have been discovered in Arabidopsis. Additionally, a key gene of production of phenylacylated-flavonols for the conferral of protection towards UV irradiation[3], was characterized by an integrative functional genomic approach. Since several physiological studies using Arabidopsis accessions have been reported with phenotypic analysis under stress conditions such as UV-B irradiation[8], drought and salinity stress[9,10] and biotic stressors[11], understanding of plant secondary metabolites for the conferral of protection towards stress condition is highly important. To capture the variance of secondary metabolites across populations, liquid chromatography-mass spectrometry (LC-MS) has often been preferred to other analytical methods as it presents the technical advantage of capturing the most extensive variety of plant metabolites.

Here, data of floral secondary metabolite abundance measured in a population of 64 *Arabidopsis thaliana* (Arabidopsis) natural accessions are presented (Data Citation 1)(Data Citation 2). Sixty-eight secondary metabolites were measured via LC-MS, ions acquired in positive and negative ion detection mode, and compounds annotated through a combination of chemical confirmation with analytical standards and comparative analysis with flavonoids knockout and over-expresser Arabidopsis lines[12,13]. The list of the Arabidopsis accessions used in this study, and raw and normalized metabolomics data are provided (Data Citation 1)(Data Citation 2), respectively. This dataset can be used for cross-study comparisons of plant metabolites, investigations on the reproducibility of metabolomics data, and in-depth analysis of plant metabolism. Importantly, transcriptomics data obtained from 10 samples in this experimental set is available in the Gene Expression Omnibus (GEO) database (Data Citation 3). Correlation studies with data of metabolomics, transcriptomics, proteomics and phenomic data of floral related traits are also anticipated. In addition, the presence in this dataset of standard reference files and complex biological data files, which were acquired on the same LC-MS system, makes it useful for practical exercises on data analysis and interpretation. Finally, as several secondary compounds initially identified in model plants bring nutritional and health benefits to humans[14,15], these data will be helpful in the design of future plant metabolic engineering used for translational genomics applications from model species to crops.

Methods

Plant material and sample preparation

Seeds of Arabidopsis natural accessions (Table 1 (available online only)) were germinated on 1/2 MS salts solidified with 1% of agar in a growth chamber (16 h light, 140-160 µmol m^{-2} s^{-1}, 20 °C; 8 h dark, 16 °C) after vernalization (two days in the dark at 8–10 °C). Fourteen days after planting, the seedlings were transferred onto soil (GS-90 Einheitserde; Gebrueder Patzer) and grown in a greenhouse (16 h light, an average irradiance of 120 µmol m^{-2} s^{-1}, 20 °C; 8 h dark, 16 °C) until flowering. Positioning of the plants was randomized during plant growth. Fully open mature flowers (first flowers) were harvested at around noon (after approximately 5 h of light) and immediately frozen in liquid nitrogen for further analysis. Flowers from three plants were individually harvested to prepare one biological replicate. Sample preparation and extraction were performed as previously described[3].

LC-MS analysis and flavonoid mutant-based peak annotation

Profiling of secondary metabolites was performed as previously described[3,16]. Briefly, flower tissues were ground with liquid nitrogen and homogenized in a mixer mill for 3 min at 25 *Hz* with a zirconia bead and 20 µL of extraction buffer (80% methanol, prepared with 5 µg mL^{-1} isovitexin as an internal standard) per mg of ground tissue (e.g., 204.0 µl extraction buffer for 10.2 mg fresh weight sample). Thereafter, the supernatant was separated from the cellular debris via centrifugation at 12,000 x G and 3 µL of the clarified supernatant directly injected in an HPLC system Surveyor (Thermo Finnigan, USA) coupled to LTQ-XP system (Thermo Finnigan, USA) for metabolite profiling described as below. All samples including flower extracts obtained from Arabidopsis mutants described in 'Data processing and metabolite data analysis' were analyzed together. Sample run order was determined by replicates consecutively.

Chromatography

Chromatography was performed as previously described[16]. Samples were run on a Surveyor HPLC system (Thermo, USA), 150 × 2 mm, 2.0 µm particles (Reverse Phase Luna $C18_{(2)}$, Phenomenex, USA), HPLC column at 28 °C oven temperature. The solvents used for the assay consisted of water containing 0.1% *v*/*v* formic acid (Solvent A) and an acetonitrile solution containing 0.1% *v*/*v* formic acid (Solvent B). Gradient [time (min)/%B] starting: 2.0/0, 4.0/15, 14.0/32, 19.0/50, 19.01/100, 21.0/100, 21.01/0, 23.0/0 at flow rate 0.20 mL min^{-1}. Injection volume was 2 µL.

Mass spectrometry

The compounds were detected using a Thermo LTQ-XL Linear-Ion-Trap mass spectrometer (expected resolution is 0.3 u FWHM) with electrospray ionization (ESI) mode in negative (collision energy: 0 and 30 meV) and positive ion detections with a scan range from 100–2000 *m/z*. Main MS parameters (capillary temperature: 275 °C; source voltage: 4.00 kV(negative) and 4.50 kV (positive); capillary voltage: −50 V(negative) and 50 V(positive) were optimized for the detection of plant secondary metabolism. Other MS parameters are described in Tohge *et al.*, 2010[16]. The LTQ-XP used the Xcalibur software (Thermo Finnigan, USA) version 2.1.0 for data acquisition.

Data processing and metabolite data analysis

Data were processed using Xcalibur 2.1.0 software, and peak identification and annotation implemented through a combination of the following approaches: standard chemical confirmation[17], MS fragmentation and retention time profiling, mutant analysis[3,12,13], literature/database survey[18,19]. The following Arabidopsis mutants known for having altered flavonoid profiling were used as control lines for the determination of flavonoid derivatives: *UDP-glucosyl transferase 78D2* (*ugt78d2*), decreased production of flavonoid-3-*O*-glucoside[20]; *transparent testa 7 (tt7)*, no production of quercetin and isorhamnetin derivatives[21]; *ugt78d1,* no production of flavonol-3-*O*-rhamnosides[22]; *ugt78d3*, no production of flavonol-3-*O*-arabinosides[23]; *O-methyltransferase 1 (omt1),* no production of isorhamnetin-derivatives[12]; *ugt89c1*, no production of flavonol-7-*O*-rhamnosides[24]; *tt4*, no production of all flavonoids[25,26]; *production of anthocyanin pigment 1-Dominant (pap1-D)*, increased accumulation of anthocyanins[20,27]. Peak picking was performed by Xcalibur Quan Browser (Window (sec), 30; highest peak; minimum peak height (S/N), 3.0; Baseline window, 80-150; area noise factor, 2; peak noise factor, 10; peak height (%), 5.0, tailing factor, 1.5).

Transcriptomic data

Transcriptomic analysis was performed using ATH1 microarrays as described previously[3] with ten accessions (Col-0, C24, Cvi, Da, Rsch, Ler-0, Ws, Sap, Stw and RLD). Duplicate hybridizations were carried out for Col-0 and C24, and a single hybridization was performed for all the other accessions except Col-0 and C24. Data is deposited in the Gene Expression Omnibus database (Data Citation 3).

Data Records

Raw data obtained from the analysis of natural Arabidopsis accessions and mutant reference lines have been deposited in the Metabolights (Data Citation 1). Raw data contains two negative (collision energy: 0 and 30 meV) and one positive ion detections. Cdf files contain negative and positive ion detections without data of in-source fragmentation using collision energy. This dataset contains a total of 216 raw files resulting from 72 lines (64 accessions and 8 Arabidopsis mutant lines) with three biological replicates each. A dataset of floral secondary metabolite (68 compounds; 16 glucosinolates, 3 hydroxycinnamate derivatives, 42 flavonol derivatives and 7 putative polyamines) and general statistics relative to the natural accessions used in the study is provided (Data Citation 2). Metabolite data was obtained from a dataset previously published[3] and reformatted for correlation-based analysis by average-scaling and log-transformation ($[\log_2(\text{mean(replicates)}/\text{mean(mean of all accessions)})]$) (Data Citation 2). The geographic coordinates of the Arabidopsis accessions provided in Table 1 (available online only) are updated accordingly with the Arabidopsis 1001 genome database (http://1001genomes.org/)[28].

Technical Validation

To qualitatively and quantitatively validate metabolite data obtained from three biological samples the standard deviation was estimated (Data Citation 2).

Usage notes

Data of floral secondary metabolites are presented in Excel files (Data Citation 2). For each compound, the method used for peak identification/annotation, which includes retention time, ion detection mode and relative peak area, is specified. The value of the relative peak area was obtained from the average of three measurements ($n = 3$) normalized by the standard deviation (SD)(Data Citation 2). Compound's family name and reference literature are also provided. The abundance of floral metabolites, normalized by average-scaling (mean/average) and log-transformation ($\log_2$) is reported (Data Citation 2). The dataset here presented can be used for cross correlation studies to integrate metabolomics with transcriptomics, proteomics, and floral phenotypic data. Figure 1 shows an example of metabolite-metabolite correlation network analysis (r^2>0.6, Pearson correlation estimated R statistical package (https://www.r-project.org/)) performed with the data reported (Data Citation 2). Visualization of network connection based on coefficient value was performed with Cytoscape (http://www.cytoscape.org/) using an organic layout style (Data Citation 2). As previously discussed[3] accession-specific floral phenylacyl-flavonol glycosides (saiginols, indicated with the number 1 in Fig. 1) show a strong correlation within the saiginol clade. The following ten additional clades of compounds were also identified and these are indicated in Fig. 1 with the following numbering: 2) common flavonol mono- or di-glycosides, 3) pollen specific flavonols and pollen specific polyamines, 4) putative pollen specific polyphenolic polyamines, 5) flavonol-3-*O*-(2′′-*O*-rhamnosyl)glucoside-7-*O*-rhamnosides, 6) flower specific flavonol-glycosides, 7) accession-specific glucosinolate, 8) short-chain aliphatic glucosinolates, 9), long-chain

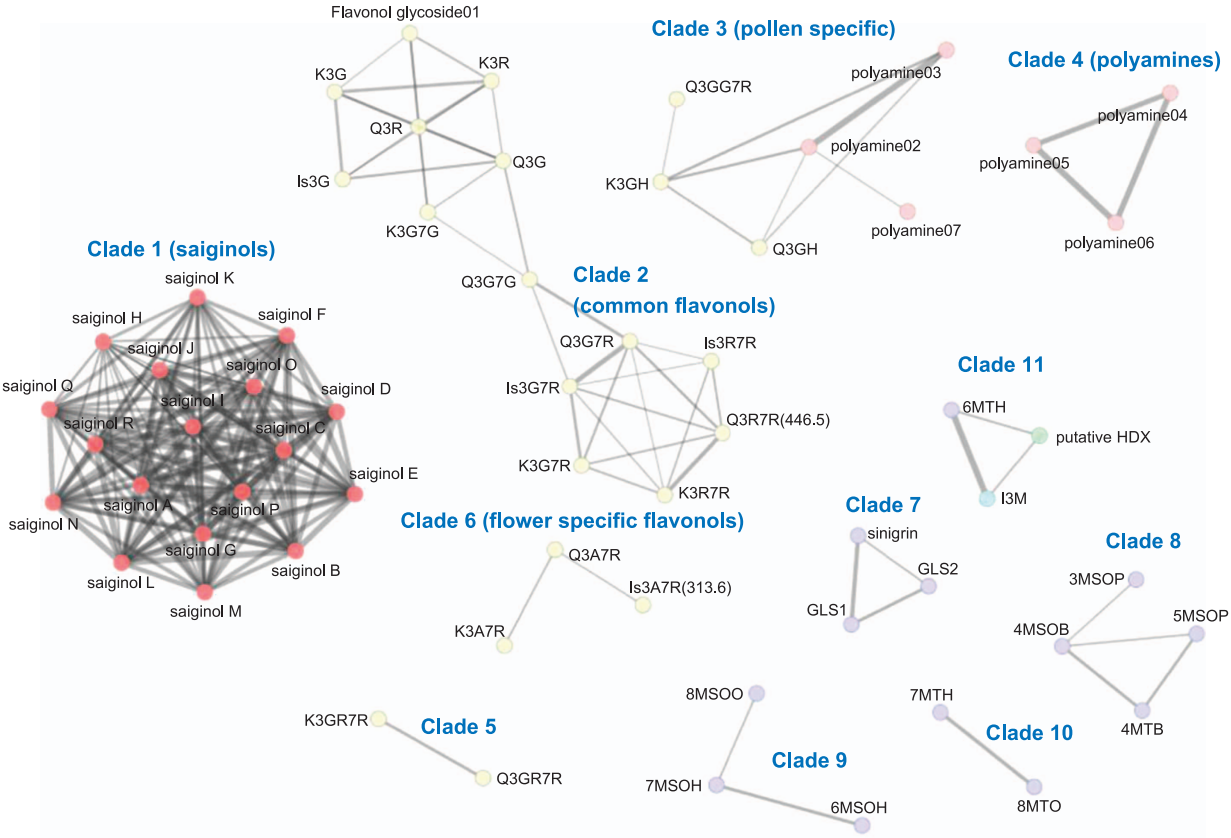

Figure 1. Correlation network of Arabidopsis floral secondary metabolites. Network analysis and visualization were performed with Cytoscape using an organic layout. The Pearson correlation threshold of 0.6 was chosen to determine the connections between edges and nodes. Nodes represent metabolites and the edges the interaction between metabolites. The size of nodes and edges maps to clustering coefficient and correlation coefficient, respectively, with small nodes and thin edges representing small values. Different classes of metabolites are represented with different colors: saiginols, red; flavonols, yellow; polyamine, pink; purple, aliphatic glucosinolates; green, putative hydroxycinnamate; light blue, indole glucosinolate.

aliphatic sulfinyl-glucosinolates, 10) long-chain aliphatic thio-glucosinolates, and 11) other glucosinolates as for example indolic glucosinolates. No subclades of hydroxycinnamates were found. Network analysis suggests that metabolites that belong to the same clade are produced in Arabidopsis natural accessions that share the common genetic polymorphism, transcriptionally co-regulated, or are the resulted of a similar metabolic pattern maintained by the combination of different metabolic flux changes. The data presented in this article are useful in biodiversity studies, e.g., to investigate relationships between natural metabolic diversity and accession distribution, physiological diversity and the genomic polymorphism.

References

1. Kliebenstein, D. J. *et al.* Genetic control of natural variation in Arabidopsis glucosinolate accumulation. *Plant Physiol.* **126,** 811–825 (2001).
2. Fernie, A. R. & Tohge, T. The genetics of plant metabolism. *Annu. Rev. Genet* **51,** 287–310 (2017).
3. Tohge, T. *et al.* Characterization of a recently evolved flavonol-phenylacyltransferase gene provides signatures of natural light selection in Brassicaceae. *Nat Commun* **7,** 12399 (2016).
4. Tohge, T. *et al.* The evolution of phenylpropanoid metabolism in the green lineage. *Crit Rev Biochem Mol Biol.* **48,** 123–152 (2013).
5. Tohge, T. & Fernie, A. R. Leveraging Natural Variance towards Enhanced Understanding of Phytochemical Sunscreens. *Trends Plant Sci.* **22,** 308–315 (2017).
6. Kliebenstein, D. J. *et al.* (2001) Gene duplication in the diversification of secondary metabolism: tandem 2-oxoglutarate-dependent dioxygenases control glucosinolate biosynthesis in Arabidopsis. *Plant Cell.* **13,** 681–6932001.
7. Ishihara, H. *et al.* Natural variation in flavonol accumulation in Arabidopsis is determined by the flavonol glucosyltransferase BGLU6. *J Exp Bot.* **67,** 1505–1517 (2016).
8. Piofczyk, T., Jeena, G. & Pecinka, A. Arabidopsis thaliana natural variation reveals connections between UV radiation stress and plant pathogen-like defense responses. *Plant Physiol Biochem.* **93,** 34–43 (2015).
9. Des Marais, D. L. *et al.* Physiological genomics of response to soil drying in diverse Arabidopsis accessions. *Plant Cell.* **24,** 893–914 (2012).
10. Bac-Molenaar, J. A. *et al.* Genome-wide association mapping of time-dependent growth responses to moderate drought stress in Arabidopsis. *Plant Cell Environ.* **39,** 88–102 (2015).

11. Ariga, H. *et al.* NLR locus-mediated trade-off between abiotic and biotic stress adaptation in Arabidopsis. *Nat Plants* **3,** 17072 (2017).
12. Tohge, T. *et al.* Phytochemical genomics in Arabidopsis thaliana: A case study for functional identification of flavonoid biosynthesis genes. *Pure and Applied Chemistry* **79,** 811–823 (2007).
13. Tohge, T., Scossa, F. & Fernie, A. R. Integrative approaches to enhance understanding of plant metabolic pathway structure and regulation. *Plant Physiol.* **163,** 1499–1511 (2015).
14. Martin, C. *et al.* Plants, diet, and health. *Annu Rev Plant Biol.* **64,** 19–46 (2013).
15. Tohge, T. & Fernie, A. R. An overview of compounds derived from the shikimate and phenylpropanoid pathways and their medicinal importance. *Mini Rev Med Chem* **17,** 1013–1027 (2016).
16. Tohge, T. *et al.* Combining genetic diversity, informatics and metabolomics to facilitate annotation of plant gene function. *Nat Protoc* **5,** 1210–1227 (2010).
17. Nakabayashi, R. *et al.* Metabolomics-oriented isolation and structure elucidation of 37 compounds including two new anthocyanins from Arabidopsis thaliana. *Phytochem* **70,** 1017–1029 (2009).
18. Tohge, T. & Fernie, A. R. Web-based resources for mass-spectrometry-based metabolomics: A user's guide. *Phytochem* **70,** 450–456 (2009).
19. de Souza, L. P. *et al.* From chromatogram to analyte to metabolite. How to pick horses for courses from the massive web-resources for mass spectral plant metabolomics. *GigaScience* **6,** 1–20 (2017).
20. Tohge, T. *et al.* Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing a MYB transcription factor. *Plant J.* **42,** 218–235 (2005).
21. Koornneef, M. *et al.* A gene controlling flavonoid-3'-hydroxylation in Arabidopsis. *Arabidopsis lnformation Service* **19,** 113–115 (1982).
22. Jones, P. *et al.* UGT73C6 and UGT78D1, glycosyltransferases involved in flavonol glycoside biosynthesis in Arabidopsis thaliana. *J Biol Chem.* **278,** 43910–43918 (2003).
23. Yonekura-Sakakibara, K. *et al.* Comprehensive flavonol profiling and transcriptome coexpression analysis leading to decoding gene–metabolite correlations in Arabidopsis. *Plant Cell* **20,** 2160–2176 (2008).
24. Yonekura-Sakakibara, K. *et al.* Identification of a flavonol 7-O-rhamnosyltranserase gene determining flavonoid pattern in Arabidopsis by transcriptome coexpression and reverse genetics. *J Biol Chem.* **282,** 14932–14941 (2007).
25. Koornneef, M. The complex syndrome of TTG mutants. *Arabidopsis lnformation Service* **18,** 45–51 (1981).
26. Jackson, J. A. *et al.* Isolation of Arabidopsis mutants altered in the light-regulation of chalcone synthase gene expression using a transgenic screening approach. *Plant J.* **8,** 369–380 (1995).
27. Borevitz, J. O. *et al.* Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. *Plant Cell.* **12,** 2383–2394 (2000).
28. *1001 Genomes Consortium*. 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. *Cell* **166,** 481–491 (2016).

Data Citations

1. Tohge, T. *MetaboLights* MTBLS528 (2017).
2. Tohge, T, Borghi, M & Fernie, A. *Figshare* http://doi.org/10.6084/m9.figshare.c.3938875 (2018).
3. *Gene Expression Omnibus* GSE83291 (2016).

Acknowledgements

T.T. and A.R.F. were funded by the Max Planck Society. MB is supported by a Marie Skłodowska-Curie Actions Individual Fellowship Grant no. 656918.

Author Contributions

T.T. and M.B. prepared tables and performed correlation network analysis. T.T., M.B. and A.R.F. wrote the manuscript.

Additional information

Table 1 is only available in the online version of this paper.

Competing interests: The authors declare no competing interests.

How to cite this article: Tohge, T, et al. The natural variance of the Arabidopsis floral secondary metabolites. *Sci. Data* 5:180051 doi: 10.1038/sdata.2018.51 (2018).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.