Molecular Plant
Research Article

CellPress
Partner Journal

Exploiting Natural Variation in Tomato to Define Pathway Structure and Metabolic Regulation of Fruit Polyphenolics in the Lycopersicum Complex

Takayuki Tohge[1,2], Federico Scossa[1,3], Regina Wendenburg[1], Pierre Frasse[4], Ilse Balbo[1], Mutsumi Watanabe[1,2], Saleh Alseekh[1,5], Sagar Sudam Jadhav[1], Jay C. Delfin[2], Marc Lohse[1], Patrick Giavalisco[1,6], Bjoern Usadel[1,7], Youjun Zhang[1,5], Jie Luo[8], Mondher Bouzayen[4] and Alisdair R. Fernie[1,5,*]

[1]Max-Planck-Institut für Molekulare Pflanzenphysiologie, Am Muehlenberg 1, 14476 Potsdam-Golm, Germany

[2]Graduate School of Biological Science, Nara Institute of Science and Technology, Ikoma 630-0192 Japan

[3]Council for Agricultural Research and Economics (CREA), Research Centre for Genomics and Bioinformatics, via Ardeatina 546 00178 Rome, Italy

[4]Université de Toulouse, INP-ENSA Toulouse, Génomique et Biotechnologie des Fruits, Castanet-Tolosan 31326, France

[5]Institute of Plant Systems Biology, 4000 Plovdiv, Bulgaria

[6]Max Planck Institute for Biology of Ageing, Joseph Stelzmann Strasse 9b, 50931 Cologne, Germany

[7]Institute of Botany and Molecular Genetics, BioSC, RWTH Aachen University, 52056 Aachen, Germany

[8]National Key Laboratory of Crop Genetic Improvement and National Center of Plant Gene Research (Wuhan), Huazhong Agricultural University, Wuhan 430070, China

***Correspondence: Alisdair R. Fernie (ernie@mpimp-golm.mpg.de)**

https://doi.org/10.1016/j.molp.2020.04.004

ABSTRACT

While the structures of plant primary metabolic pathways are generally well defined and highly conserved across species, those defining specialized metabolism are less well characterized and more highly variable across species. In this study, we investigated polyphenolic metabolism in the lycopersicum complex by characterizing the underlying biosynthetic and decorative reactions that constitute the metabolic network of polyphenols across eight different species of tomato. For this purpose, GC–MS- and LC–MS-based metabolomics of different tissues of *Solanum lycopersicum and wild tomato species* were carried out, in concert with the evaluation of cross-hybridized microarray data for MapMan-based transcriptomic analysis, and publicly available RNA-sequencing data for annotation of biosynthetic genes. The combined data were used to compile species-specific metabolic networks of polyphenolic metabolism, allowing the establishment of an entire pan-species biosynthetic framework as well as annotation of the functions of decoration enzymes involved in the formation of metabolic diversity of the flavonoid pathway. The combined results are discussed in the context of the current understanding of tomato flavonol biosynthesis as well as a global view of metabolic shifts during fruit ripening. Our results provide an example as to how large-scale biology approaches can be used for the definition and refinement of large specialized metabolism pathways.

Key words: *Solanum lycopersicum*, secondary metabolism, natural diversity, wild accessions, pathway elucidation, gene discovery

Tohge T., Scossa F., Wendenburg R., Frasse P., Balbo I., Watanabe M., Alseekh S., Jadhav S.S., Delfin J.C., Lohse M., Giavalisco P., Usadel B., Zhang Y., Luo J., Bouzayen M., and Fernie A.R. (2020). Exploiting Natural Variation in Tomato to Define Pathway Structure and Metabolic Regulation of Fruit Polyphenolics in the Lycopersicum Complex. Mol. Plant. **13**, 1027–1046.

INTRODUCTION

Crop domestication and the genetic bottleneck that it tends to create has led to a massive decrease in the allelic diversity of

Published by the Molecular Plant Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and IPPE, CAS.

the gene pools of modern cultivars (Tanksley and McCouch, 1997; Fernie et al., 2006). As such, natural genetic resources provide a good source of exotic germplasm for crop-breeding strategies (Zamir, 2001; McCouch, 2004; McCouch et al., 2013). The development and relative cheapness of next-generation sequencing (Schneeberger and Weigel, 2011) and genome-wide association mapping (Platt et al., 2010; Tian et al., 2011; Fernie and Gutierrez-Marcos, 2019) have, however, led to increasing adoption of this technique for the characterization of other complex traits including metabolism and growth (Keurentjes et al., 2006; Schauer and Fernie, 2006; Riedelsheimer et al., 2012; Bellucci et al., 2014; Zhu et al., 2018). Given that the nutritional and calorific value of crops are, by and large, determined by their chemical composition, identifying the genetic bases that control the accumulation of metabolites is of fundamental importance for attempts at crop improvement. For this reason, many studies have been carried out utilizing natural variation to study both metabolite accumulation and metabolic regulation (Kliebenstein, 2009; Sulpice et al., 2010; Chan et al., 2011; Wen et al., 2014; Ishihara et al., 2016; Tohge et al., 2016; Perez de Souza et al., 2019).

Tomato is a powerful crop model given the availability of a wealth of genome information (Tomato Genome Consortium, 2012; 100 Tomato Genome Sequencing Consortium, 2014; Bolger et al., 2014; Lin et al., 2014). Moreover, its diploid nature renders its genetics relatively facile (Klee and Giovannoni, 2011). Furthermore, the lycopersicum complex consists of the cultivated species *Solanum lycopersicum* and over 10 wild species that it can be crossed with (albeit not all of these crosses produce self-fertile offspring; Covey et al., 2011), with high-quality genome-sequence information also being available for *Solanum pennellii* (Bolger et al., 2014) and *Solanum pimpinellifolium* (Tomato Genome Consortium, 2012). Genome-sequence information is also available for *Solanum arcanum* and *Solanum habrochaites* (100 Tomato Genome Sequencing Consortium, 2014). Tomato is additionally arguably the best-characterized crop species at the metabolomic level with several surveys of primary metabolites, vitamins and antioxidants, volatiles and volatile precursors, cuticular waxes, and specialized metabolites being carried out in tomato populations subject to a wide range of genetic or environmental interventions (Tikunov et al., 2005, 2013; Schauer et al., 2008; Ballester et al., 2010; Nashilevitz et al., 2010; Schilmiller et al., 2010; Chan et al., 2011; Dal Cin et al., 2011; Falara et al., 2011; Yeats et al., 2012b; Itkin et al., 2013; Schwahn et al., 2014; Alseekh et al., 2015, 2017; Tieman et al., 2017; Zhu et al., 2018). Among these studies relatively few, however, have surveyed a wide range of species from the lycopersicum complex. Schauer et al. (2005) documented the levels of primary metabolites in leaf and fruit tissue of *S. lycopersicum* and five further members of the complex, highlighting a considerable but not massive variance in the levels of primary metabolites, most notably those documented to play a role in stress adaptation. Yeats et al. (2012a) looked at cuticular wax of a total of seven species and demonstrated that such screens can be used to understand the ecological and evolutionary functional genomics (Mitchell-Olds et al., 2008) of metabolic pathways. Similarly, Iijima et al. (2013) and Schwahn et al. (2014) studied glycoalkaloid contents across the lycopersicum complex and used the data obtained to refine the pathway structure of this crucially important class of metabolites. More recently, Zhu et al. (2018) revealed a multi-omics view of tomato metabolism by utilizing the breeding history of tomato in the study of metabolic polymorphism between tomato cultivars as a method of forward genetic screening. Despite the paucity of studies of the parental species themselves, several studies of primary metabolism and volatile emissions using populations resulting from crosses between *S. lycopersicum* and *S. pennellii* and *S. habrochaites* and *Solanum chmielewskii* have been carried out (Schauer et al., 2006, 2008; Tieman et al., 2006; Stevens et al., 2007; Do et al., 2010; Dal Cin et al., 2011; Steinhauser et al., 2011; Perez-Fons et al., 2014), as have similar studies in *Arabidopsis* (Ishihara et al., 2016; Tohge et al., 2016), rice (Chen et al., 2016; Okazaki and Saito, 2016), and maize (Riedelsheimer et al., 2012; Wen et al., 2014). Furthermore, considerable research effort has been invested in understanding the genetics of specialized metabolism (Keurentjes et al., 2006; Rowe et al., 2008; Fernie and Tohge, 2017). Dissection of the genetic architecture of specialized metabolism in *Arabidopsis* and tomato has been particulary successful, with both quantitative trait loci and genome-wide association studies proving successful at pinpointing the genetic control of metaboliteaccumulation (Keurentjes et al., 2006; Angelovici et al., 2013; Alseekh et al., 2015; Tieman et al., 2017; Ye et al., 2017; Wu et al., 2018; Fernie and Gutierrez-Marcos, 2019).

In recent years, tomato has become a predominant model for investigating flavonoids and phenylpropanoids (Dal Cin et al., 2011; Zhang et al., 2015). The reasons for this are manifold; these compounds play roles in many important processes *in planta* including pigmentation of fruits and vegetables, plant–pathogen interactions, and protection against high-light, high-salt, and chilling conditions, and serve as precursors for volatile production (Tieman et al., 2006). Moreover, these polyphenolic compounds are an integral part of the diet, and there are increasing reports that dietary polyphenols are likely candidates for the observed beneficial effects of a diet rich in fruits and vegetables on the prevention of cardiovascular diseases and other chronic diseases including diabetes and obesity (Martin et al., 2011; Tohge and Fernie, 2017).

There is a long history of identification of tomato flavonoids; for example, rutin (quercetin-3-*O*-(6″-*O*-rhamnosyl)glucoside, Q3Glc6″Rha) was identified in 1924 in the leaves of cv. Ailsa Craig (Charaux, 1924). Subsequently, 10 flavonol-glycosides were identified by chromatographic co-elution with standard compounds (Wu and Burrell, 1958; Moco et al., 2006; Iijima et al., 2008; Mintz-Oron et al., 2008; Slimestad et al., 2008; Slimestad and Verheul, 2011; Shahaf et al., 2016). The two most abundant tomato anthocyanins, nasunin and petanin, were characterized (Tohge et al., 2015) in *Del/Ros1* transgenic purple tomato (Butelli et al., 2008). Dihydrochalcone and naringenin derivatives, such as naringenin chalcone (NC) and its glucoside, which are common fruit-specific stilbenoids, were identified in tomato fruit peel over 60 years ago (Wu and Burrell, 1958; Iijima et al., 2008), while dihydrochalcone glycoside, phloretin-3′5′-di-*C*-glucoside (Phe3′Glc5′Glc), was found in tomato peel much more recently (Slimestad et al., 2008). Seven chlorogenic acid (CGA)-related compounds, chlorogenate (3-CGA),

cryptochlorogenate (4-CGA), neochlorogenate (5-CGA), and three di-caffeoyl-type and one tri-caffeoyl-type chlorogenates, have been found in tomato tissues (Luo et al., 2008; Shahaf et al., 2016). Furthermore, the CGA derivatives caffeoyl-2-*O*-glucarate (Caf2Glr) and caffeoyl-5-*O*-glucarate (Caf5Glr) were identified as vegetative green tissue-specific metabolites in tomato (Teutschbein et al., 2010). Other hydroxycinnamate derivatives such as sinapoyl-1-*O*-glucoside (Sin1Glc) and feruloyl-1-*O*-glycoside (Fer1Glc) have additionally been characterized in tomato tissues (Shahaf et al., 2016) Moreover, phenolamide (*N*-caffeoyl-putrescine), which is a member of the metabolite family of Solanaceae-conserved conjugates of hydroxycinnamate and polyamine, has long been known to be present in tomato leaf extracts (Martin-Tanguy et al., 1978). Some polyphenolic compounds, such as NC and its glycosides, have been found to be metabolic markers for fruit ripening in model tomato cultivars (Carrari et al., 2006; Rohrmann et al., 2011). However, metabolic shifts of polyphenolics as well as primary metabolite precursor pools during fruit ripening in tomato species are still not well investigated, although such shifts need to be elucidated for the development of a metabolomics-assisted breeding approach employing introgression between domesticated cultivars and wild species.

In this study, we performed a cross-species survey of polyphenolic content in five different tissues of *S. lycopersicum* and a further seven members of the *Solanum* complex (*S. pimpinellifolium*, *Solanum cheesmaniae*, *S. chmielewskii*, *Solanum neorickii*, *Solanum peruvianum*, *S. habrochaites*, and *S. pennellii*). As well as profiling the content of the polyphenolics themselves, we also evaluated the levels of their primary metabolite precursor pools. We additionally constructed a complex species-wide metabolic framework of tomato polyphenolics and successfully predicted the qualitative expression pattern of the genes involved in this framework on the basis of our metabolomic data. We subsequently assessed the transcript profiles of these genes using publicly available RNA-sequencing data and microarray data for young green fruits from all eight species to compile species-specific metabolic networks of primary and polyphenolic metabolism. In addition to 18 known genes, we identified a further eight key genes annotated as enzymatic genes involved in the flavonoid biosynthetic pathway via integration of transcriptomic data, phylogenetic analysis, and our metabolomic data. In addition, we evaluated the function of candidate glucosyltransferase genes by *in planta* and *in vitro* functional characterization with confirmation of the anticipated metabolic changes in their target phenylpropanoids, demonstrating the fidelity of this approach in pathway annotation.

RESULTS

The Current Blueprint of Polyphenolic Metabolism in *S. lycopersicum* Tissues

As a first step toward understanding the entire framework of tomato polyphenolic metabolism, we surveyed tomato polyphenols characterized by nuclear magnetic resonance studies or confirmed by co-elution with standard compounds in liquid chromatography–mass spectrometry (LC–MS) (Figure 1A). To assess tissue specificities of each metabolite class, we performed specialized metabolite analysis using ultra-performance liquid chromatography–Fourier transform mass spectrometry (UPLC–FTMS) on extracts from the immature fruit, mature fruit, leaf, stem, and root of the *S. lycopersicum* cultivar (LYCO) (M82). The major phenylpropanoids, Q3Rha6″Glc, K3Rha6″Glc, Q3Glc, K3Glc, NC, and seven CGAs, could be identified by co-elution with standard compounds. In addition, the tomato peel extracts used in Iijima et al. (2008) (MicroTom, National Bioresource Project Tomato, MEXT, Japan) and Moco et al. (2007) (M82), were analyzed via the KomicMarket (http://webs2.kazusa.or.jp/komicmarket/) or Moto (http://www.ab.wur.nl/moto/) databases, as were *Nicotiana tabacum* leaves (Niggeweg et al., 2004; Luo et al., 2008; Ruprecht et al., 2016) as an aid in peak annotation. Caffeoyl-glucarate, which is a 3CGA derivative, was annotated using its mono-isotopic accurate mass ($C_{15}H_{16}O_{11}$, isotopic molecular weight 372.069265), which allowed us to separate it from the mass peak of 5OH-feruloyl-glucosides ($C_{16}H_{20}O_{10}$, isotopic molecular weight 372.105649). This allowed us to confirm earlier reports that this metabolite accumulates in a leaf-specific manner (Teutschbein et al., 2010). However, since two peaks were observed, they were annotated as caffeoyl-glucarate isomers. Some hydroxycinnamate derivatives were also confirmed on the basis of elution profiles matching those of *Arabidopsis* flower extracts (Tohge et al., 2016).

Flavonol triglycosides were detected at high levels in all tissues except roots, but high levels of flavonol monoglycosides were detected only in fruit peels, leaves, and stems. Chalcones, such as NC and P3′Glc5′Glc, were, however, only detected in peel extracts and mature fruits (Figure 1B). Furthermore, despite the high accumulation of mono-CGAs in the leaf and stem in comparison with fruit tissues, di- and tri-CGAs were only observed in mature fruit tissue samples. Hydroxycinnamate derivatives were observed in all tissues. Given the tissue specificities of major polyphenols of each different biosynthetic branch (Figure 1C), we decided to chemically profile these five tissue types in detail. The tissue specificity of these known polyphenolic compounds was subsequently employed for the analysis of the natural variation of tomato accessions and the reconstruction of the polyphenolic biosynthetic framework of the lycopersicum complex.

Metabolomic Signatures Defining Fruit Ripening in Tomato Accessions

To gain a global overview of the biosynthetic framework of tomato polyphenolics, including tissue/species-specific biosynthetic branches, we chose seven major tomato species (*S. pimpinellifolium*, PIMP; *S. cheesmaniae*, CHEE; *S. chmielewskii*, CHMI; *S. neorickii*, NEOR; *S. peruvianum*, PERU; *S. habrochaites*, HABR; and *S. pennellii*, PENN) for our experiments (Supplemental Table 1 and Figure 2A). Whole fruits were harvested at immature and mature stages according to the ripening indicators described in Grumet et al. (1981) (Supplemental Table 1 and Figure 2A). Although several phenylpropanoid metabolites are predominantly peel specific (Figure 1B), we used whole fruits for metabolomic profiling in an attempt to understand the global metabolic changes occurring during fruit ripening.

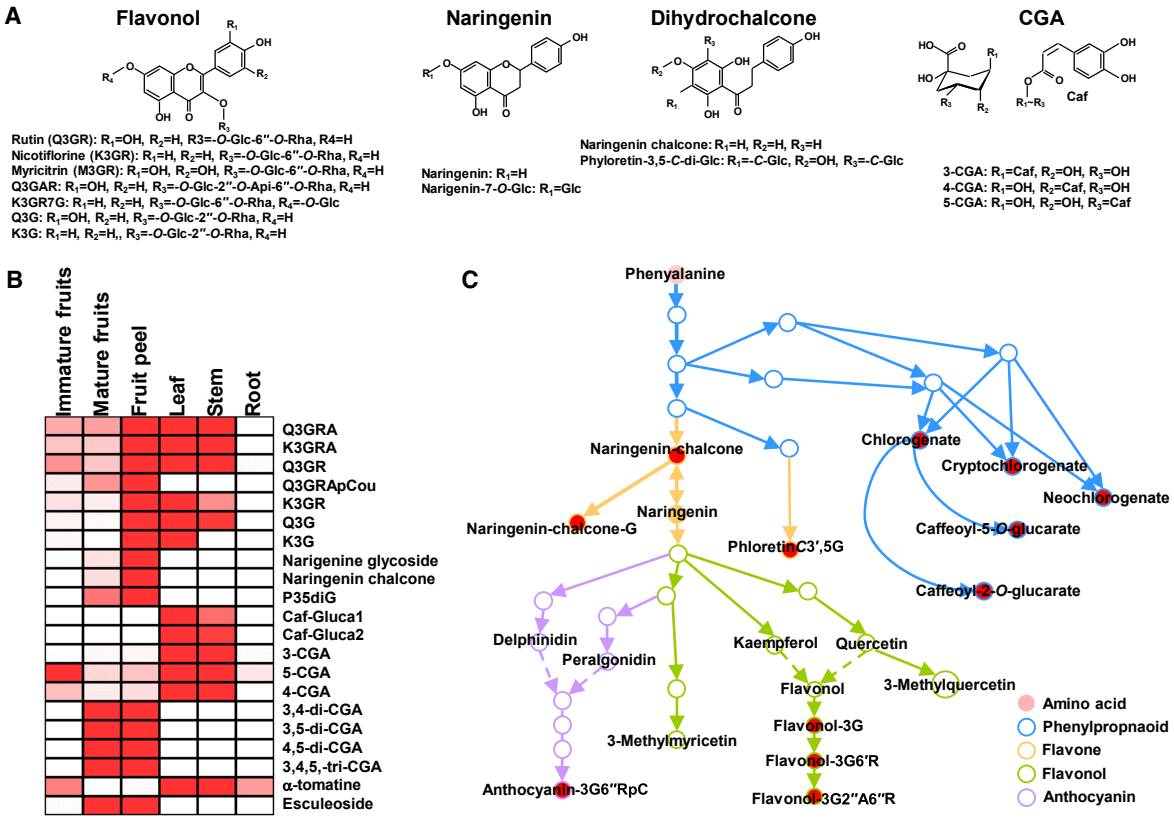


Figure 1. Tissue Specificity of Major Specialized Metabolites in M82.

(A) Major polyphenolic secondary metabolites in *S. lycopersicum*. Flavonols. naringenins, dihydrochalcones, and CGA are shown.

(B) Heatmap visualization of specialized metabolite contents analyzed by LC–MS in tomato tissues. The analysis was conducted with three independent biological replicates. Relative peak areas are visualized by a color gradient from red (high) to white (low).

(C) The known biosynthetic framework of tomato polyphenolic metabolites constructed by linking the major known polyphenolics. Colors: blue, hydroxycinnamates; orange, stilbenoids; purple, anthocyanins; green, flavonols.

Api, apiose; Caf, caffeic acid; CGA, chrologenic acid; Glc, glucose; Rha, rhamnose; 3CGA, 3-caffeoylquinate; 4CGA, cryptochlorogenate; 5CGA, neochlorogenate; Q, quercetin; G, glucose; R, rhamnose; H, hexose; A, apiose; P, phloretin; pCou, *p*-coumaric acid.

Next, to study changes in general metabolites during fruit ripening, we performed metabolite profiling using gas chromatography–mass spectrometry (GC–MS) (Lisec et al., 2006) on immature and mature fruits of the eight tomato species (Figure 2C and Supplemental Table 2). To obtain an initial classification of the mature and immature fruits based on the metabolic changes they undergo during fruit ripening, we performed hierarchical clustering analysis (HCA) on the primary metabolite dataset (Figure 2B), rendering a clear separation between immature and mature fruits with the exception of PENN fruits.

In comparison with the known metabolic changes in red-fruited species such as LYCO and PIMP, whose ripening is relatively well characterized at the metabolic level, diverse metabolic changes were observed during fruit ripening in the different species (Figure 2C). Interestingly, a decrease of β-alanine and γ-aminobutyric acid during ripening could be observed only in watery tomatoes, LYCO, PIMP, CHEE, CHMI, and NEOR, whereas the loss of threonine, glutamine, and branched-chain amino acids (BCAAs: valine, leucine, and isoleucine) was observed in all tomato accessions except PENN. By contrast, ripening-associated increases of lysine and aspartate were among the common observations made in all species. Furthermore, there was a negative correlation between serine and glycerate levels in all green-fruited species while quinate accumulated to remarkably high levels in the mature fruits of CHMI. In contrast to the observed decrease of malate in red/yellow carotenoid-containing fruits, green-fruited accessions such as CHMI, PERU, and PENN showed a significant increase in this organic acid during fruit ripening. Phenylalanine, which is a precursor of phenolic metabolites, decreased in all species except PENN. Intriguingly, there were very few common metabolic markers of ripening among primary metabolites across the species studied. That said, changes in both conserved and species-specific ripening markers suggest that the fruit-ripening stages of wild tomatoes, which have non-pigmented mature fruit, were properly defined in our study.

Next, we performed transcriptome analysis of fruit samples using the TOM2 microarray, focusing on the expression differences arising from genes of primary metabolism. Immature fruit samples were used for microarray analysis due to the problems caused by the differences in water contents of mature fruits between tomato species. The possibility of cross-hybridization to the array oligo probes (Supplemental Figure 1 and

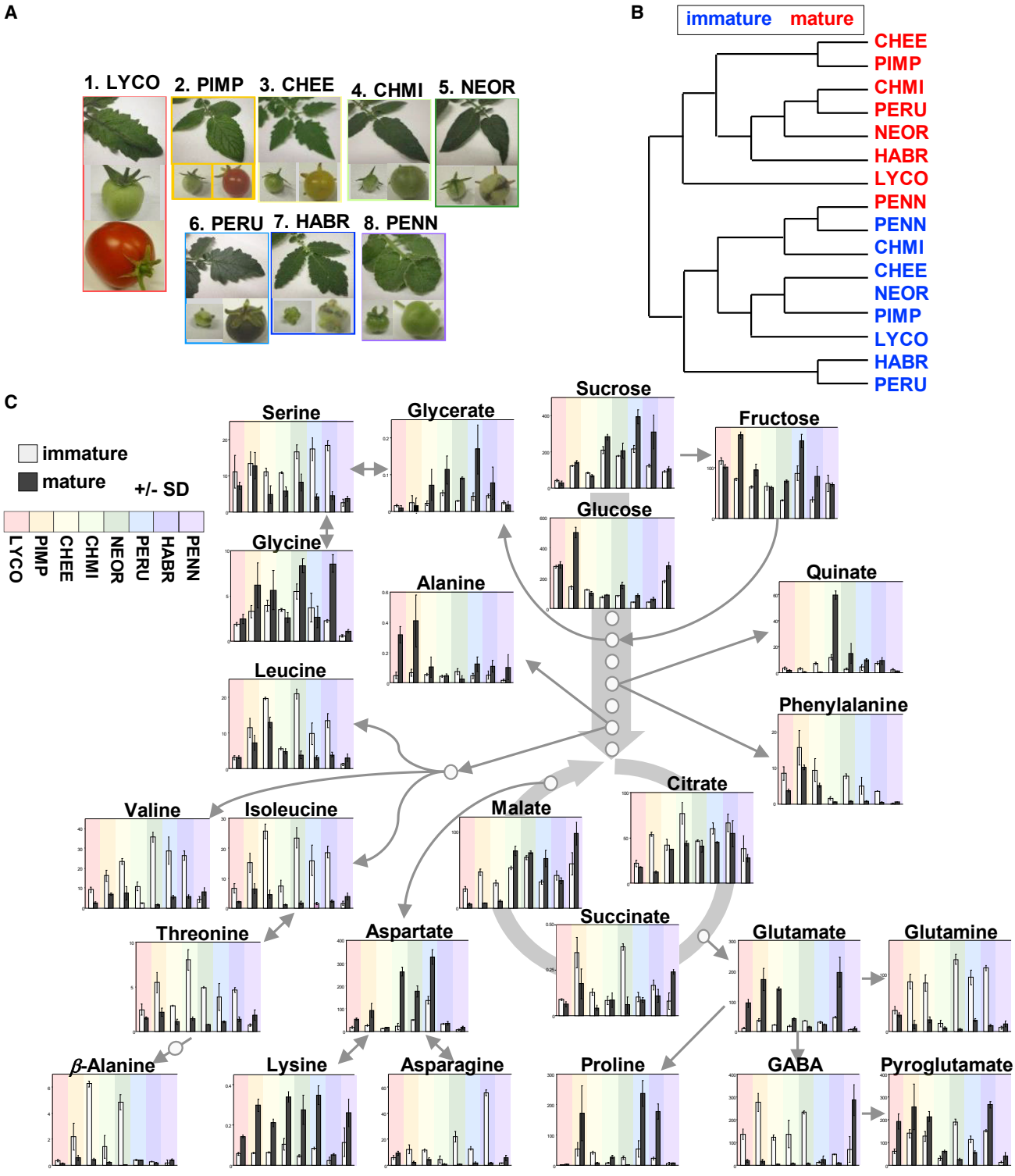


Figure 2. Metabolomic Signatures Defining Fruit Ripening in Diverse Tomato Species.

(A) Tomato species used in this study and their source of origin. 1, M82 (*S. lycopersicum*); 2, PIMP (*S. pimpinellifolium*); 3, CHEE (*S. cheesmaniae*); 4, CHMI (*S. chmielewskii*); 5, NEOR (*S. neorickii*); 6, PERU (*S. peruyianum*); 7, HABR (*S. habrochaites*); 8, PENN (*S. pennellii*). Note that LYCO is not demarcated on the map but is thought to originate from Mexico.

(B) Difference in primary metabolites analyzed by GC–MS between immature and mature fruits of different tomato accessions. Hierarchical clustering analysis was performed using MeV software (http://www.tm4.org/) with Pearson's correlation.

(C) Changes in major primary metabolites during fruit ripening. White and black colors represent mature and immature fruits, respectively.

Supplemental Table 3), whose sequences were designed from the Lycopersicon Combined Build #3 unigene database (http://www.sgn.cornell.edu), made it possible to acquire expression data from the polymorphic transcripts of the different tomato species. Finally, we compared the gene expression levels in the diverse tomato species with those in LYCO. We determined changes in gene expression at a global scale using the tomato MapMan mapping files (Urbanczyk-Wochniak et al., 2006) (Supplemental Figure 2). The similar metabolic pattern between LYCO and PIMP at the immature fruit stage was reflected in the similar expression levels of genes involved in carbon metabolism from both tomato species. Furthermore, the relatively higher levels of amino acids, with the exception of BCAA in LYCO, reflected the observation of relatively lower expression of genes involved in amino acid metabolism. The levels of tricarboxylic acid (TCA) metabolites, citrate and malate, were similar among tomato species, with transcriptomic data also revealing similar expression patterns of TCA-related genes.

Analysis of Species-Specific Metabolic Polymorphisms in Polyphenolic Metabolism

To ascertain the complete biosynthetic pathway of tomato polyphenolics, we performed metabolite profiling using LC–MS (Giavalisco et al., 2009; Tohge and Fernie, 2010) on the exact same samples used for primary metabolite profiling (Supplemental Table 4).

Three putative flavonol tetraglycosides with distinctive *m/z* of 903.24, 887.24, and 917.26 were not found in CHMI, NEOR, HABR, and PENN. Given the absence of flavonol-3Glc6″Rha7Glc in HABR and PENN, and the absence of flavonol-3Glc2″Api6″Rha in CHMI and NEOR, we suggest that these peaks correspond to downstream products of flavonol-3Glc2″Api6″Rha7Glc (quercetin/kaempferol/isorhamnetin). This is also supported by the presence of fragments corresponding to intact aglycone ions at positive ion detection (303, 287, and 317 *m/z*). In addition, due to the general enzymatic substrate specificity of flavonol-glycosyltransferases, which have a higher selectivity toward sugar donors but often share flavonol aglycones as sugar acceptors, K3Glc2″Api6″Rha and K3Glc2″Api6″Rha7Glc were found to have the expected retention time and MS spectra. Furthermore, seven putative methylated-quercetin derivatives sharing this fragment peak (317.0661 *m/z* at positive ion detection) were detected. One of the derivatives, a putative methylated-quercetin monohexoside, was confirmed as isorhamnetin(Is)-3Glc (Is3Glc) on the basis of co-elution with the peak identified in flower extracts of *Arabidopsis* Columbia-0 (Col-0) and the *atomt1* mutant (Tohge et al., 2007). Since the other six putative methylated-quercetin derivatives represent the same decorative forms of kaempferol and quercetin derivatives, these peaks were therefore annotated as isorhamnetin glycosides. In addition, the low accumulation of a putative Is3Glc6″Rha in HABR clearly correlated with the low accumulation of kaempferol/quercetin-3Glc6″Rha, reflecting the lower expression of flavonol-3Glc-6″-*O*-rhamnosyltransferase. Interestingly, six peaks were found that were only present in HABR. Analysis of MS spectra suggested that these peaks correspond to flavonol-di-hexosides and were annotated as flavonol-3Glc7Glc on the basis of co-elution with this compound from *Arabidopsis* flower extracts (Tohge et al., 2016). In addition, two further flavonol-pentosyl-hexoside peaks were annotated as flavonol-3Glc2″Api via cross-reference to the KomicMarket and MotoDB databases, while the co-elution with compounds from *Arabidopsis* flower and leaf extracts allowed 1-*O*-sinapoyl glucoside to be identified. Fifteen phenylacyl-flavonol-glycosides with *p*-coumaroyl, caffeoyl, feruloyl, and sinapoyl moieties were identified as quercetin/kaempferol derivatives, but no phenylacyl-isorhamnetin-glycosides were found. Furthermore, on the basis of peak annotation and intermetabolite correlations, additional flavonol derivatives were identified on the basis of the masses of known metabolite modifications and the known masses of the aglycones (Supplemental Table 4). This combined analysis finally resulted in the identification or putative annotation of 68 peaks, which are summarized in Supplemental Table 3.

The tissue specificity and interspecific diversity of the individual polyphenolics were evaluated by HCA using metabolite profiles obtained from leaves, immature fruits, and mature fruits of the diverse tomato species (Figure 3). Metabolite clustering by HCA resulted in clear separation between the compound families flavonol-glycosides, putative phenylacyl-flavonols, hydroxycinnamates, chlorogenate-related, and stilbenoids (Figure 3), according to their species- and tissue-specific accumulation patterns.

Despite displaying trends in the levels of quercetin and kaempferol similar to those in the other species, isorhamnetin accumulated to exceptionally high levels in HABR. Additionally, flavonol-glycosides except flavonol-7Glc were highly detected in HABR. Flavonol-tetra-glycosides accumulated to high abundance in PERU while phenylacyl-flavonols were found in leaves of LYCO, PIMP, PERU, and PENN, as well as mature fruits of LYCO and CHEE. Finally, P35diGlc was found in mature fruits of the red-fruited tomato species as well as in PERU and HABR.

Using the metabolic abundances across different tissues and/or species and the structures of the 68 identified/putatively annotated compounds, we were able to assemble a comprehensive overview of the biosynthetic framework of polyphenol metabolism in the lycopersicum complex. Figure 4 provides a summary of the resulting refined pathway. Comparison of the density of this network featuring 16 metabolites and 28 genes with that of the network in Figure 1C featuring 55 metabolites and 37 genes demonstrates the power of this approach in the refinement of pathways of plant-specialized metabolism.

Prediction of Gene Expression from Metabolomic Data

To ascertain whether it is possible to qualitatively predict the expression level of genes involved in the refined biosynthetic framework, we first compared the total amount of each metabolite related to the target enzymatic genes in each tissue and species (Supplemental Figure 3). Following this step, we predicted qualitative transcriptomic differences in the genes encoding flavonol decoration enzymes (Figure 5A and 5B) based on a presence/absence call of metabolite accumulation. Although this approach is susceptible to the potential discrepancies

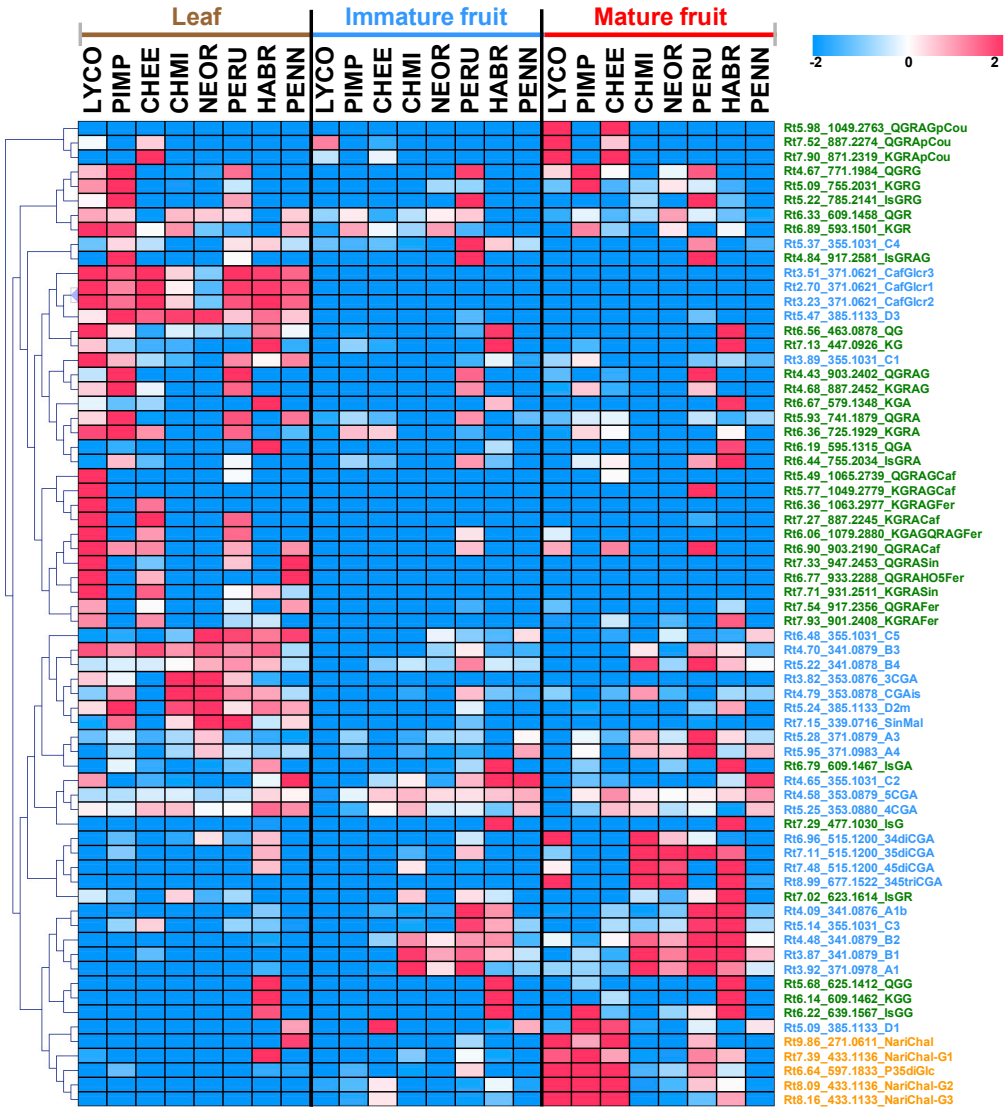


Figure 3. Metabolomic Analysis of Leaves, Immature Fruits, and Mature Fruits of Tomato Accessions by LC–MS.
Hierarchical clustering analysis was performed using MeV software with Pearson's correlation. Analysis was conducted with the average of three independent biological replicates. Relative peak area was normalized by the average value of all the tissues and all the accessions. Fold changes, in base 2 logarithmic scale, are visualized by a color gradient from red (high) to blue (low). Color: blue, hydroxycinnamates; orange, stilbenoids; green, flavonols.

between transcript and metabolite amounts (Fernie and Stitt, 2012), it can give some preliminary insights into whether a gene is active within a species- or tissue-specific flavonol biosynthetic branch. For example, given that flavonol-7-*O*-glucosides were not detected in PENN fruits (Figure 5B) and leaves (Supplemental Figure 4), the gene encoding flavonol-7-*O*-glucosyltransferase is predicted to be absent or non-expressed in PENN. In the case of flavonol derivatives, our results show that some intermediates (flavonol-3-*O*-glucoside derivatives) accumulated to high levels in both fruits and leaves of HABR, suggesting the negligible expression of genes encoding glycosyltransferases (UGTs) with flavonol-3-*O*-Glc-6-*O*-rhamnosyltransferase activity (Figure 5B and Supplemental Figure 4). Furthermore, CHMI and NEOR showed no accumulation of flavonol-3Glc2″Api in fruits and leaves, suggesting the absence or lack of expression of a gene encoding F3G2″ApiT in both species.

Cross-Species Annotation of Genes Involved in Flavonoid Biosynthesis

We next performed genome-based orthologous gene searches in order to fit genes encoding well-known enzymes to our refined polyphenolic pathway. By matching metabolite structure to enzymatic capabilities, a total of 26 enzymatic genes involved in flavonoid biosynthesis were annotated. The genes with asterisks presented in Table 1 are well-known or experimentally characterized genes, which are involved in phenylpropanoid and flavonoid biosynthesis in tomato (Tohge et al., 2015). In our survey, 10 genes, namely *SlCHS1* (*TCHS1*, CAA38980, Solyc09g09151), *SlCHS2* (*TCHS2*, CAA38981, Solyc05g053550) (O'Neill et al., 1990), *SlDFR* (Bongue-Bartelsman et al., 1994), *SlF3′5′H*, *FOMT1* and *FOMT2* (Schmidt et al., 2011, 2012), *ScAN1* (Schreiber et al., 2012), *AnthOMT* (Gomez Roldan et al., 2014), *SlCGT* (Teutschbein

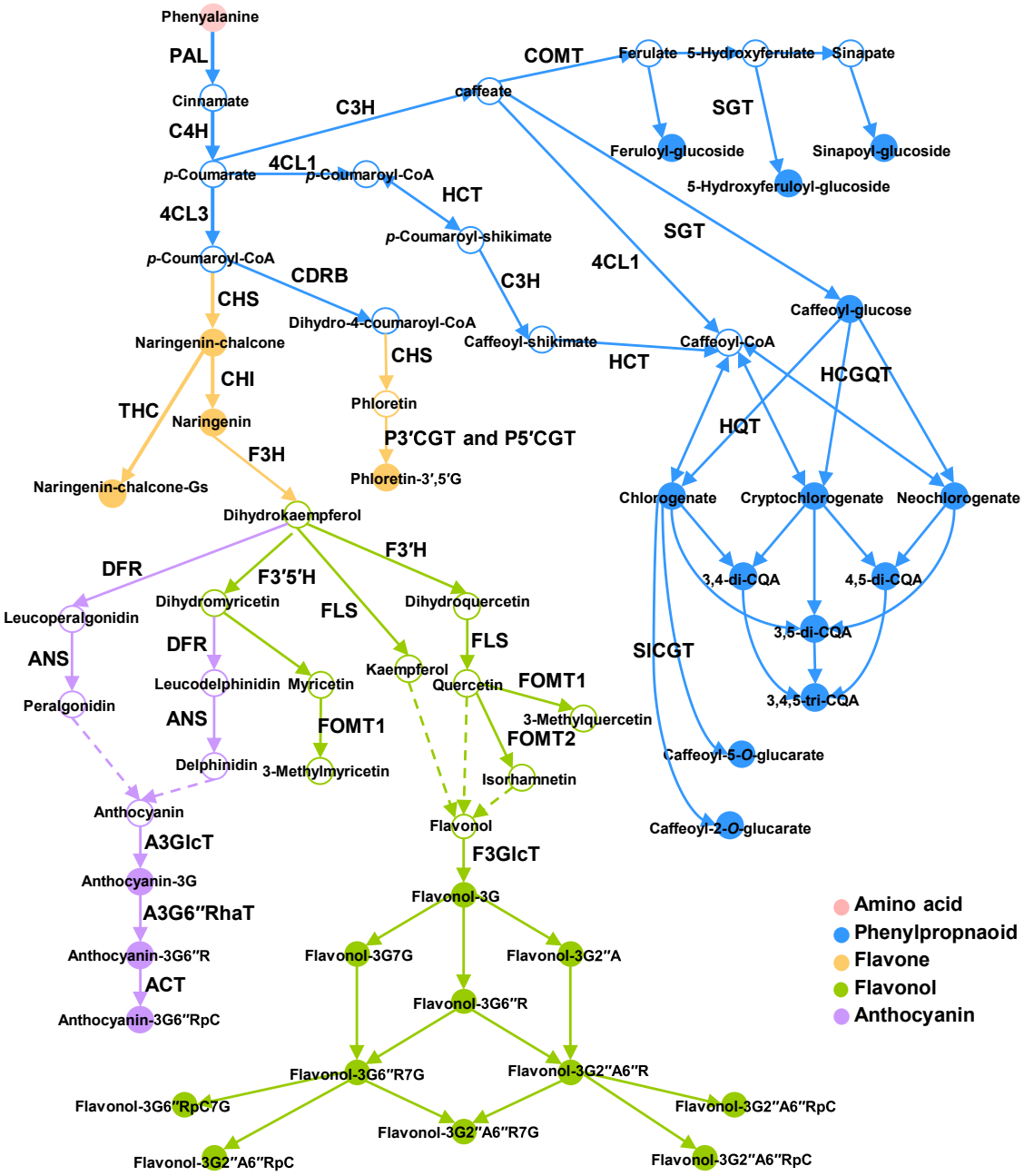


Figure 4. An Overview of the Entire Polyphenol Biosynthetic Framework of the Lycopersicum Complex.
Reconstructed polyphenol biosynthesis pathways according to the chemical reactions predicted from chemical structures of detected polyphenols in the diverse tomato species. Lines indicate chemical reaction by enzyme, and circle nodes indicate metabolite with the color of filled (detected) or empty (undetected) circles indicating the type of metabolite. PAL, phenylalanine ammonia-lyase; C4H, cinnamate-4-hydroxylase; 4CL, 4-coumarate CoA ligase; CAD, cinnamoyl-alcohol dehydrogenase; F5H, ferulate 5-hydroxylase; C3H, coumarate 3-hydroxylase; ALDH, aldehyde dehydrogenase; CCR, cinnamoyl-CoA reductase; HCT, hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyltransferase; CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase; F3′H, flavonoid-3′-hydroxylase; F3GlcT, flavonoid-3-*O*-glycosyltransferase; FOMT, flavonoid *O*-methyltransferase; FCGT, flavone-*C*-glycosyltransferase; FLS, flavonol synthase; F3GT, flavonoid-3-*O*-glycosyltransferase; DFR, dihydroflavonol reductase; ANS, anthocyanidin synthase.

et al., 2010), and *SlFdAT1* (Tohge et al., 2015), were found to have been experimentally characterized as flavonoid enzymatic genes in tomato species. The genes encoding common enzymes of phenylpropanoid biosynthesis (such as PAL, C4H, 4CL, CHI, and F3H), which are conserved in many dicot plants including *Arabidopsis*, were also annotated on the basis of sequence similarity (Table 1). Combining these approaches, we annotated 19 known genes and a further seven key genes putatively involved in the tomato flavonoid biosynthetic pathway. Seven key genes (*CDRB*, *P3′5′CGT*, *OMT1*, *Fd3GlcT*, *F3G2″ApiT*, *Fd3G6″RhaT*, and *F7GlcT*), which we initially considered to be putatively involved in the

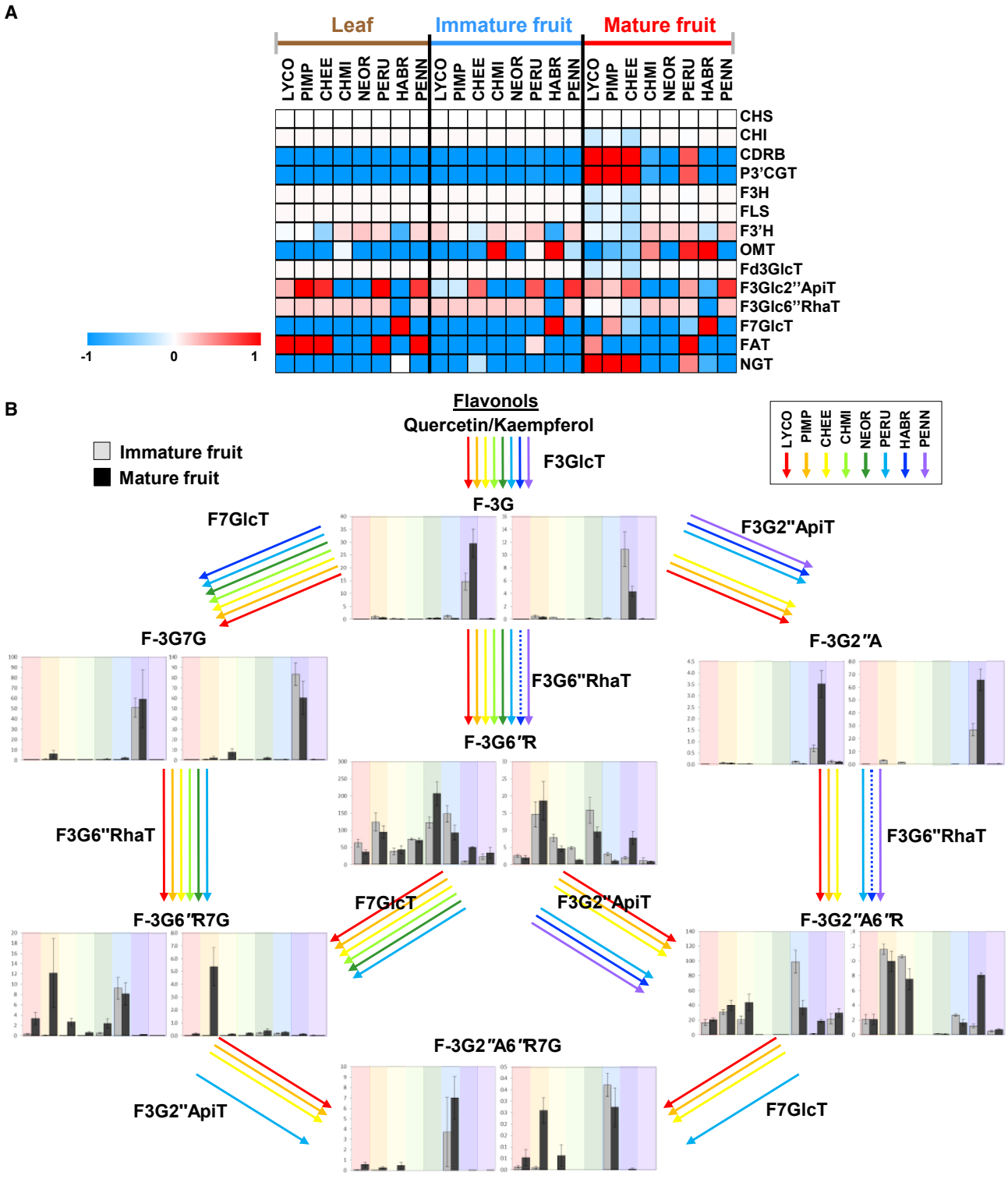


Figure 5. Proposed Relative Expression Patterns of Genes Involved in the Annotated Biosynthetic Framework.

(A) Relative expression levels of genes were estimated from the metabolomic data. Estimated gene expression levels, in the base 2 logarithmic scale of fold change, are visualized by a color gradient from red (high) to blue (low). Intensity of expression level was predicted as fold change of (sum of downstream compounds/sum of all compounds) (Supplemental Figure 1).

(B) Schematic of metabolite flow among the flavonoid decoration steps in the diverse tomato species. Gray and black indicate metabolite level in immature and mature fruits, respectively.

Gene name	Gene annotation	ITAG3.0
*CHS1**	Chalcone synthase 1	Solyc05g053550
*CHS2**	Chalcone synthase 2	Solyc09g091510
*CHI**	Chalcone isomerase	Solyc05g010320
*CHI**	Chalcone isomerase	Solyc05g010310
*CHIL**	Chalcone synthase-like	Solyc05g052240
*F3H**	Flavanone 3-hydroxylase	Solyc02g083860
*F3′H**	Flavonoid 3′-hydroxylase	Solyc03g115220
*F3′5′H**	Flavonoid 3′5′-hydroxylase	Solyc11g066580
*FLS1**	Flavonol synthase	Solyc11g013110
CDRB	*p*-Coumaroyl-CoA reductase	Solyc10g078740
P3′5′CGT	Phloretin-3′,5′-C-glucosyltransferase	Solyc02g088500
*F3′H**	Flavonoid 3′-hydroxylase	Solyc03g115220
OMT1	*O*-Methyltransferase	Solyc03g080180
Fd3GlcT	Flavonoid-3-*O*-glucosyltransferase	Solyc10g083440
F3G2″ApiT	Flavonol-3-*O*-glc-2″-*O*-apiosyltransferase	Solyc10g008860
Fd3G6″RhaT	Flavonoid-3-*O*-glc-6″-*O*-rhamnosyltransferase	Solyc09g059170
F7GlcT	Flavonoid-7-*O*-glucosyltransferase	Solyc10g079350
*FOMT1**	Flavonol-*O*-methyltransferase 1	Solyc06g083450
*FOMT2**	Flavonol-*O*-methyltransferase 2	Solyc06g007960
*FOMT3**	Flavonol-*O*-methyltransferase 3	Solyc06g064500
*DFR**	Dihydrokaempferol-4-reductase	Solyc02g085020
*ANS**	Anthocyanin synthase	Solyc08g080040
*AnthOMT**	Anthocyanin-*O*-methyltransferase	Solyc09g082660
*A5GlcT**	Anthocyanin-5-*O*-glucosyltransferase	Solyc12g098590
*SlFdAT1**	Flavonoid-3-*O*-rutinoside-4‴-phenylacyltransferase	Solyc12g088170
*SlCGT**	GDSL lipase-like caffeoyltransferase	Solyc01g099020

Table 1. Annotation of Tomato Genes Involved in Secondary Metabolism.
Orthologous genes were found using a combined approach based on Sol BLAST searches using the protein sequences of already characterized *Arabidopsis*, tobacco, and apple genes as queries.
*Genes characterized or annotated previously.

polyphenol pathway, have now been placed in the flavonoid biosynthesis pathway as orthologs of other known genes catalyzing the same enzymatic reactions in other plant species. To extend our annotation to the enzymatic reactions predicted by our assembled framework, we used known genes encoding enzymes with reactions similar to those of the target biosynthetic branch, such as *p*-coumaroyl-CoA reductase (MdCDRB, apple) (Dare et al., 2013), phloretin-3,5-*C*-di-glucosyltransferase (FcCGT, citrus) (Ito et al., 2017), chalcone-4-*O*-glucosyltransferase (Chl4′GlcT, carnation and snapdragon) (Ono et al., 2006), phloretin-2-*O*-glucosyltranferase (Phl2′GlcT, MdPGT1F, apple [Jugdé et al., 2008] and DicGT4, carnation [Ogata et al., 2004]), and flavone-*C*-glucosyltransferase (OsF6CGlcT, rice) (Brazier-Hicks et al., 2009), as bait in a search for tomato orthologs using ITAG3.0 (Table 1). We obtained orthologs of CDRB and CGT genes in tomato, SlCDRBs (Soly10g078740) and P3′5′CGT (Solyc02g088500), respectively (Table 1). This compiled list was next used for integrative analysis of metabolomic and transcriptomic data.

Integration with Genomics and *In Silico* Transcriptomic Data

As a first approach, we surveyed publicly available RNA sequence data of tomato species and cross-referenced them with our metabolic data in order to predict transcriptionally active (metabolic) genes in tomato. In this regard, we used gene expression data obtained from LYCO and PENN fruits (Bolger et al., 2014) as well as leaves of different tomato species (Koenig et al., 2013), because some differences in polyphenolic levels were observed in both fruits and leaves. Figure 6A shows *in silico* expression analysis of flavonoid biosynthetic genes in leaves of four tomato species (LYCO, PIMP, HABR, and PENN) (Koenig et al., 2013) and mature fruit of two tomato accessions

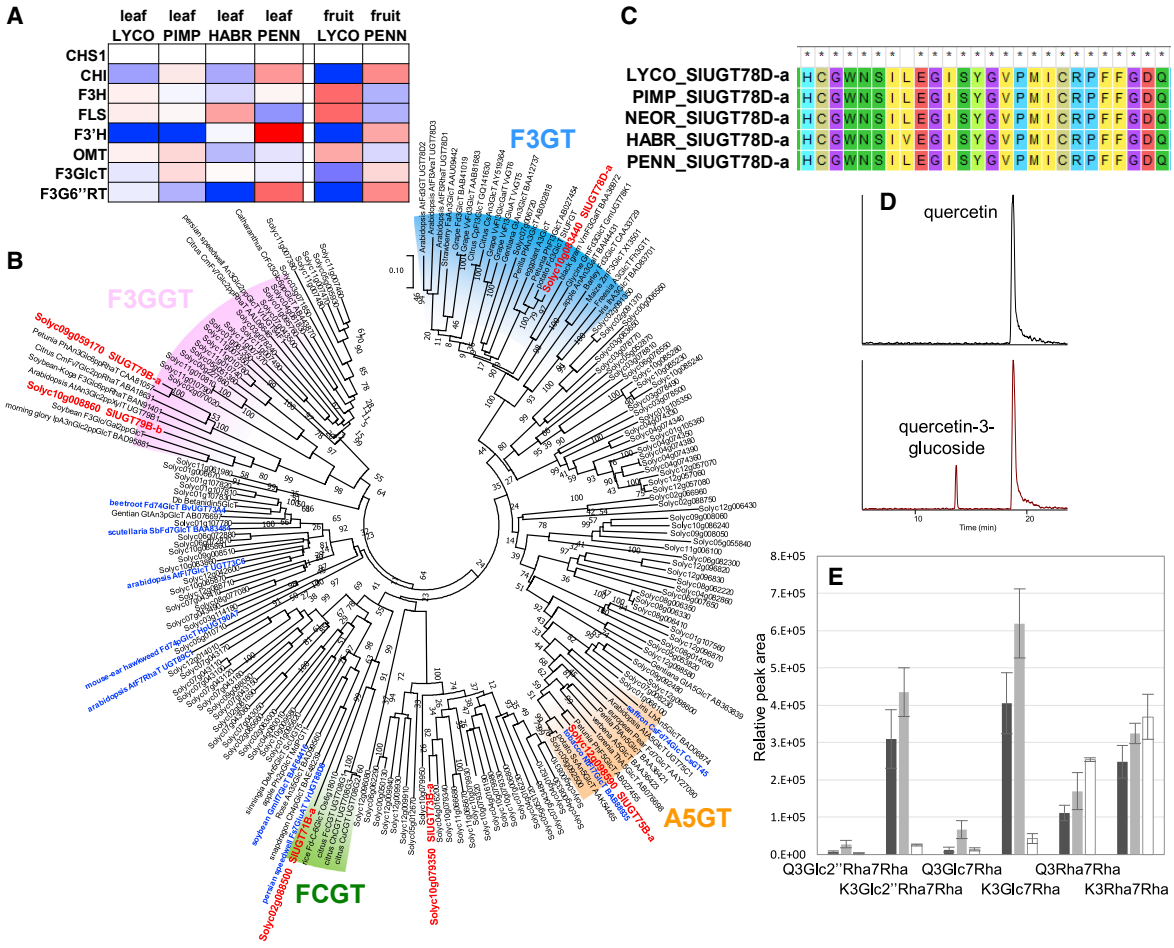


Figure 6. Prediction of Functional UGT Genes Involved in the Tomato Flavonoid Biosynthetic Pathway.

(A and B) (A) *In silico* expression analysis of common flavonoid biosynthesis genes. **(B)** Molecular phylogenetic tree of the amino acid sequences of all tomato UGT genes. A phylogenetic tree was constructed with the aligned UGT1 protein sequences in MEGA 8.0 (http://www.megasoftware.net/) using the neighbor-joining method with the following parameters: Poisson correction, complete deletion, and bootstrapping (1000 replicates, random seed). Red indicates candidate UGTs annotated in the main text. Blue indicates flavonoid 7-*O*-glycosylatransferase.

(C) Comparison of UGT PSPG sequence between LYCO, PIMP, NEOR, HABR, and PENN.

(D and E) (D) Assay of recombinant Solyc10g083440 protein. **(E)** Profiling of flavonol-glycosides in Solyc10g083440-overexpressing transgenic *Arabidopsis*. Three to six biological replicates were used for the analysis. Error bars indicate SE. Red and blue indicate increase and decrease, respectively. Q, quercetin; NariChal, naringenin chalcone; Phil, phloretin; Glc, glucose; Rha, rhamnose.

(LYCO and PENN) (Bolger et al., 2014). We first evaluated *in silico* expression data with predicted gene expression patterns by cross-species metabolomic analysis. F3′H, which is a key enzymatic gene involved in the production of kaempferol or quercetin, shows a higher level of accumulation in PENN in comparison with that in LYCO, PIMP, and HABR. Additionally, expression of the gene encoding F3G6″RhaT was clearly lower in HABR and higher in PENN (Figure 6A). Although we observed different magnitudes of differential expression between the *in silico* transcriptome data and our predictions, overall the results suggest that metabolite profiling data, when showing distinct metabolic phenotypes, can be used for the prediction of gene expression within a comprehensive metabolic network.

To assess candidate genes involved in the flavonoid glycosylation steps, we compared candidate flavonoid UGT genes putatively involved in tomato flavonoid decoration by phylogenetic tree analysis with characterized genes encoding flavonoid UDP-glycosyltransferases from other plant species (Figure 6B). This phylogenetic analysis using all tomato UGT genes resulted in a clear separation of three major clades, (1) F3GT, flavonoid-3-*O*-glycosyltransferases, (2) A5GT, anthocyanin-5-*O*-glycosyltransferases, and (3) FGGT, flavonoid-3Glc-glycosyltransferases. Despite the clear separation of three clades according to the sugar attachment position, F7GT (flavonoid-7-*O*-glycosyltransferases) did not exhibit a clear separation (Figure 6B). With the classification of the UGT subclade, candidate UGT genes for each glycosylation step were annotated (Supplemental Figure 5).

The candidate gene of flavonoid glucosyltransferase, *SlUGT78D-a* (Solyc10g083440), was also annotated as anthocyanin 3-*O*-glucosyltranferase in an analysis of purple transgenic tomato (Tohge et al., 2015). We further sequenced and analyzed Solyc10g083440 of LYCO, PIMP, NEOR, HABR, and PENN (Figure 6C). The active PSPG sites of UDP-

glycosyltransferases obtained from the other tomato species are highly conserved. This result supported the conserved role of this UGT in the 3-*O*-glycosylation step among different tomato species. Another candidate gene that was highly expressed in HABR leaves, *SlUGT79B-b* (Solyc10g008860), clustered close to the clade of flavonoid-3Glc-2″-*O*-glucosyltransferases (Figure 6B). Its expression pattern and phylogenetic proximity are thus consistent with the further decoration of flavonol-3-*O*-Glc, namely the step catalyzed by flavonol-3Glc-2″-*O*-apiosyltransferases in tomato species. By contrast, another candidate gene (*SlUGT71B-a*, Solyc10g079350), which is putatively involved in the decoration of flavonol-7-*O*-glucosides in HABR seedlings based on our expression data and proposed pathway structure (Supplemental Figure 5), was not clustered in a distinct subclade of flavonoid-glycosyltransferases. The fact that flavonoid-7-*O*-glycosyltransferases were distributed to several clades could be interpreted as evidence that the region specificity of the glycosylation in position 7 emerged multiple times during the radiation of glycosyltransferases. For example, while all three *Arabidopsis* flavonoid-3-*O*-glycosyltransferases (flavonoid-3-*O*-glucosyltransferase AtUGT78D2; flavonol-3-*O*-rhamnosyltransferase, AtUGT78D1; flavonol-3-*O*-arabinosyltransferase AtUGT78D3) were classified as AtUGT78D subfamily members, flavonol-7-*O*-glucosyltransferase and flavonol-7-*O*-rhamnosyltransferase, which show lower amino acid sequence similarity, were instead classified into different subfamilies (AtUGT73C6 and AtUGT89C1, respectively). These results suggest that SlUGT73B-a (Solyc10g079350) might be involved in the flavonoid-7-*O*-glucosyltransferase reaction in tomato. The enzyme encoded by the candidate UGT gene named *SlUGT71B-a* (Solyc02g088500) was separated from the other *O*-glycosyltransferases and clustered close to flavonoid-*C*-6-glucosyltransferase in rice (Brazier-Hicks et al., 2009) and citrus (Ito et al., 2017). This result suggests that SlUGT71B-a acts as a phloretin-3′,5′-*C*-di-glucosyltransferase (P3′5′CGT). A comprehensive summary of the annotations of candidate genes within tomato polyphenolic biosynthesis pathways is shown in Table 1.

Functional Validation of Candidates of Flavonoid-Glycosyltransferase in Tomato Species

We chose to confirm the activity of UGTs using a transformation-based approach. To validate the function of candidate genes annotated as flavonol-glycosyltransferases, we cloned the cDNA of SlSolyc10g083440 (*SlUGT78D1-a*, annotated as 3-*O*-glucosyltransferase) from LYCO. Additionally, genetic polymorphism among tomato species was investigated using *SlUGT78D1-a*, genes obtained from LYCO, PIMP, NEOR, HABR, and PENN (Figure 6C). The function of SlUGT78D1-a was verified either by assessing the metabolite profiles of *Arabidopsis* and tomato-overexpressing lines or tomato virus-induced gene silencing (VIGS) lines in which SlUGT78D1-a expression was suppressed. Firstly, we tested SlUGT78D1-a activity in a recombinant protein assay (Figure 6D). Furthermore, SlUGT78D1-a was transformed into *Arabidopsis thaliana* (Col-0), which normally accumulates flavonoid-3Glc. In comparison with wild type and a knockout mutant of *Arabidopsis* for AtF3GlcT (At5g17050, AtUGT78D2) (Tohge et al., 2005), displaying lower accumulation of flavonol-3Glc, *A. thaliana plants* overexpressing SlUGT78D1-a showed higher accumulation of flavonol-3Glc-related compounds (Figure 6D). We next silenced candidate genes in tomato fruit (LYCO, MicroTom) using an established VIGS system (Orzaez et al., 2009; Alseekh et al., 2015). Following agroinjection into unripe fruit, ripe fruits were harvested at 10 days after the breaker stage. Flavonol contents of methanolic extracts of fruit pericarp samples were evaluated by LC–MS analysis. Flavonols in fruits from the silenced lines were compared with those in the fruits from plants infiltrated with a pTRV2 empty vector. The levels of flavonol-3-*O*-glucosides were clearly lower in the VIGS suppression lines of SlSolyc10g083440, which we previously annotated as a SlF3GlcT, and higher in the transiently overexpressing lines (Supplemental Figure 6). In the lines transiently overexpressing SlUGT73B-a, which is a candidate flavonol-7-*O*-glucosyltransferase, the level of Q3Glc6″Rha7Glc was relatively low. However, since the sequence of SlUGT73B-a is not similar to the sequences of proteins in other F7GT clades, further characterization of this protein is required.

DISCUSSION

In this study, we conducted a cross-species analysis of the chemodiversity in secondary metabolites among several tissues of tomato species with the aim of understanding how this diversity emerged and is structured across the lycopersicum complex. The reconstruction of interspecific and intertissue diversity of secondary metabolites, and also the annotation/identification of the candidate metabolic genes, has elucidated a "blueprint" of the pan-species biosynthetic pathway of polyphenols. Tomato species represent an excellent system for unraveling the diversity in this pathway, considering that the tomato clade includes a set of species with a wealth of genomic and metabolomic resources available (see Tohge and Fernie, 2015), which diverged recently (with an estimated clade age of 2–7 million years ago [Haak et al., 2014]) but nevertheless show considerable diversity at the morphological and physiological levels (Peralta et al., 2008).

Domestication of tomato has clearly affected the phenotypic variation in the lycopersicum complex, as the result of conscious selection of wild ancestors with preferred phenotypes for agricultural uses: during this process, the global allelic diversity has been narrowed by selective pressures centered mainly on fruit yield and productivity (Schauer et al., 2005; Koenig et al., 2013; Zhu et al., 2018). Recently, to introduce some of the typical domesticated phenotypes into a wild genome background, alleles for tomato fruit yield and nutritional value were introduced to *S. pimpinellifolium* or modified by a genome-editing approach, creating a *de novo* domesticated form of a wild species that retained many of its ancestral phenotypes related to disease resistance and stress tolerance (Li et al., 2018a; Zsögön et al., 2018). These studies indicate that tomato species began to separate long before human intervention, reflecting the interactions of fruits with non-human seed dispersers, with human involvement being important for only one small part of the lycopersicum complex. At a wider level, domesticated tomato cultivars have been studied in comparison with their wild relatives in order to gauge the full extent of their phenotypic diversity and understand the genetic basis of domestication and diversification traits (Fernie and Tohge, 2017; Zhu et al.,

2018). Although the large majority of these studies—in cereals as well as in fruit crops—has focused on the genetic dissection of "classical" traits of the domestication syndrome (e.g., fruit/seed size [Chakrabarti et al., 2013; Zuo and Li, 2014], seed dormancy [Wang et al., 2018], inflorescence/plant architecture [Boden et al., 2015], photoperiod response [Müller et al., 2016, 2018]), metabolic traits have also been shown recently to be direct targets of selection.

In fact, metabolic comparisons between domesticated and wild tomato have often revealed that wild tomatoes produce a much higher level and broader diversity of metabolites such as amino acids (Schauer et al., 2005), polyphenolics (Alseekh et al., 2015; Zhu et al., 2018), glycoalkaloids (Iijima et al., 2008; Zhu et al., 2018), and acyl-sugars (Schilmiller et al., 2010). This is most likely explained by the fact that wild tomatoes have undergone adaptation to their habitats in a manner that includes the production of phytochemical protectants such as specialized compounds. Therefore, metabolomics-assisted breeding via species-wide metabolomic comparison currently emerges as a key approach that can be used for metabolic crop improvements resulting in additional production of stress protectants such as polyphenolics (Schauer et al., 2005; Nakabayashi et al., 2014; Tohge et al., 2015). Indeed, these metabolic improvements can be obtained with minimal targeted modifications of metabolic loci by genome-editing approaches, thus limiting the penalties associated with the strong selection for "classical" domestication traits (e.g., fruit size), which often affected negatively the content of fruit metabolites (Tieman et al., 2017; Zhu et al., 2018). Furthermore, tomato polyphenolics, for example flavonoids and chlorogenic acids, have been well studied for their pleiotropic health-beneficial effects (Butelli et al., 2008; Scarano et al., 2017; Tohge and Fernie, 2017; Li et al., 2018a, 2018b). This is likely one of the reasons why the metabolic diversity of polyphenolics in the fruit pericarp and glandular trichomes of *S*. *pennellii*, *S*. *pimpinellifolium*, and *S*. *habrochaites* has been much investigated (Schauer et al., 2005; Schilmiller et al., 2010; Fan et al., 2016, 2017). However, it is important to bear in mind that flavonoids and chlorogenic acids also play important functions in the plant cell itself, with critical roles being reported in tolerance to UV, cold, and drought (Clé et al., 2008; Schulz et al., 2016; Davies et al., 2018) as well as conferral of resistance to a wide number of diseases and pests (Niggeweg et al., 2004; Zhang et al., 2013). Despite these powerful motivations to study the flavonoid content of crops, there are as yet surprisingly few genomic-based studies of natural variation in this class of compounds. Structural elucidation of flavonoids is in fact complex: although the general biosynthesis of the aglycones has been elucidated, multiple biochemical routes exist in plants for their decoration, leading to a large diversification of this family across different species or even between different tissues of a single plant (Tohge et al., 2013). Given the existence of lineage-specific pathways for flavonoid biosynthesis, we have adopted here a cross-species metabolomic analysis using diverse tissues of eight tomato species to reveal the entire biosynthetic framework of this class of metabolites in the lycopersicum complex. A total of 68 polyphenolic metabolites were detected in this analysis, which is considerably higher than the 17 known at the inception of this work. There are vast limitations given the low availability of commercial standards; however, making use of a combinatorial phytochemical approach including validation between literature- and web-based resources (Moco et al., 2006; Iijima et al., 2008; Tohge et al., 2014), and the use of biological reference extracts to aid in the identification of detected peaks, led to considerable advances being made in our ability to uncover the pan-species network of polyphenols in the lycopersicum complex. Additionally, the utilization of high-resolution MS allowed the discrimination of the accurate chemical formulas of several compounds, for example caffeoyl-glucarate ($C_{15}H_{16}O_{11}$) and 5OH-feruloyl-glucosides ($C_{16}H_{20}O_{11}$), which produce isobaric molecular ions, and determination of their tissue-specific accumulation patterns (Teutschbein et al., 2010). Such combinatorial approaches, especially when applied in a phylogenetic framework, allow reconstruction of the structural diversity of various classes of secondary metabolites and reconstruction of entire lineage-specific pathways (Schwahn et al., 2014; Brockington et al., 2015; Moghe et al., 2017). Such a strategy represents a highly efficient method by which it is possible to link experimental data to those of species-specific warehouse databases (Afendi et al., 2012) and to the broader scientific literature; as such, this approach represents an important route by which the identification of the vast number of unknown metabolites in current metabolomics research can be tackled. This will be particularly useful for specialized metabolism in which peak annotation of known and previously reported compounds on the basis of co-elution with reference extracts is highly valuable (Iijima et al., 2008; Tohge et al., 2011).

As mentioned above, the complexity of peak annotation of secondary metabolites is a common bottleneck for full pathway elucidation and for subsequent functional validation of candidate genes. To define pathway structure and metabolic regulation of (fruit) polyphenolics in the lycopersicum complex, we used a combination of qualitative and quantitative metabolite profiling data. In *Arabidopsis*, such qualitative and quantitative metabolite information regarding the variation between tissues, as well as the analysis of knockout mutants of genes involved in secondary metabolism, revealed the whole framework of polyphenolic metabolism (Routaboul et al., 2006; Yonekura-Sakakibara et al., 2007; Stracke et al., 2009; Saito et al., 2013). Given that knockout mutants of polyphenolic metabolism in tomato are currently not available, metabolic data obtained from the analysis of natural variation between tomato species and tissues is the only way to reconstruct metabolic pathways across entire phylogenetic lineages; this approach was also used here for the estimation of putative gene expression patterns (Figure 5A).

Expression patterns of key genes estimated by metabolic flux on the refined biosynthetic framework was predicted by combining accurate metabolomic data with data of species and tissue specificities. Generally, accumulation of major polyphenolics is controlled neither by metabolite transport nor degradation systems. Therefore, this approach can be employed for any plant species including non-model plants. Indeed it offers the following advantages over other experimental approaches: (1) lower complexity of sequence assembly following ortholog identification; (2) cheaper cost; (3) facile collection of dynamic data; (4) information concerning the complexity of convergent evolution of gene functions. Finally, prediction of expression patterns of possible enzymatic genes was integrated with transcriptomic

data for gene annotation. This approach thus provided a case study for a novel strategy of metabolomics-assisted functional genomics of key genes involved in plant polyphenolic pathways. Since a comprehensive overview of plant polyphenolic biosynthesis is incomplete, combinatorial strategies involving the integration of metabolomic and gene expression/functional data in a phylogenetic context are highly useful for studies of plant metabolism. This is particularly true for plant-specialized metabolism, and given our considerable and increasing reliance on natural products, it seems likely that such approaches will greatly empower our capability to engineer the levels of such compounds both in the species in which they naturally occur and in other appropriate plant and microbial species.

In the current study, we identified and annotated some of the key genes involved in the biosynthesis of wild tomato-specific flavonoids (Alseekh et al., 2015). These results will provide the foundation for the analysis of their *in planta* biological functions against environmental stress. In the analysis of floral secondary metabolism in *Arabidopsis*, phenylacylated flavonols, which were exclusively produced by the accessions originating from southern Europe and Africa, and hence likely adapted to regions of high light intensity, were found to be the key compounds related to metabolic adaptation under UV-B light stress (Tohge et al., 2016). In our metabolite analysis, a total of 67 polyphenolics were found in tomato species. The major polyphenols previously known in domesticated tomato fruits showed much lower accumulation as well as considerably less diversity with respect to wild tomato fruits. The genes driving the diversification of polyphenol synthesis found in this study could be used for target approaches in metabolomics-assisted breeding. As such, the work described here considerably expands upon the previous analysis of metabolic variation within the LYCO × PENN introgression lines (Alseekh et al., 2015; Fernie and Tohge, 2017).

In summary, we elucidated here the whole biosynthetic framework of polyphenol biosynthesis across eight species of the lycopersicum complex, with detailed annotation of metabolites and structural genes. Given the large chemical diversity of polyphenolic compounds in tomato species, and their multiple physiological roles conferring beneficial traits, the results will likely be useful for metabolomics-assisted breeding approaches and integrative -omics approaches with further high-resolution transcriptomic data, such as those emanating from RNA-sequencing and analysis of genetic polymorphisms. Furthermore, the metabolic changes during fruit ripening in tomato species also represent one of the key targets for the design of future genetics-based metabolic engineering. In addition, these approaches could be applied to any important plant species to define the function of specific metabolites (Alseekh and Fernie, 2018; Alseekh et al., 2017), as well as to extract the maximal biological knowledge from metabolomic data.

METHODS

Plant Materials and Cultivation

In addition to the crop model tomato (*S. lycopersicum*, LYCO), seven major wild tomato species (*S. pimpinellifolium*, PIMP; *S. cheesmaniae*, CHEE; *S. chmielewskii*, CHMI; *S. neorickii*, NEOR; *S. peruvianum*, PERU; *S. habrochaites*, HABR; *S. pennellii*, PENN) were used this study (Supplemental Table 1). Three species (LYCO, PERU, and PIMP) were described by Linnaeus (1753). From a taxonomical perspective, nine tomato species are defined as the major tomato species (Child, 1990; Peralta and Spooner, 2000). However, given that *S. chilense* is a self-incompatible wild species, we could not obtain enough fruits under our growth conditions and this species was, therefore, omitted from our study. Tomato seeds of the accessions LA3475 (LYCO), LA1589 (PIMP), LA0428 (CHEE), LA1028 (CHMI), LA2133 (NEOR), LA1274 (PERU), LA1777 (HABR), and LA0716 (PENN) were obtained from the true-breeding monogenic stocks maintained by the Tomato Genetics Stock Center (University of California, Davis). The seeds were germinated on Murashige and Skoog medium containing 2% (w/v) sucrose and were grown in a growth chamber at 500 λmol photons m^{-2} s^{-1} and 25°C under a 12/12-h light/dark regime. Experiments were carried out on mature fully expanded source leaves from the plants 10 days after germination. Mature fruits of tomato accessions were harvested according to the ripening indicator described in Grumet et al. (1981). *A. thaliana* Col-0, *tt4* mutant (Shikazono et al., 1998), *f3gt* (Tohge et al., 2005), *f3at* (Yonekura-Sakakibara et al., 2008), and *atomt1* (Yonekura-Sakakibara et al., 2007) plants were grown on conventional soil under standard greenhouse conditions. Leaves of tobacco (*N. tabacum*) and maize (*Zea mays*) were obtained from plants cultivated on conventional soil under standard greenhouse conditions. Plant materials were harvested and immediately frozen with liquid nitrogen. Tomato peel extracts were kindly provided by Dr. Koh Aoki.

UPLC–FTMS Analysis of Specialized Metabolites

UPLC separation of specialized metabolites was performed according to a previously published protocol (Giavalisco et al., 2009). In brief, a Waters Acquity UPLC system (Waters, Milford, MA, USA), equipped with an HSS T3 C18 reversed-phase column (100 × 2.1 mm internal diameter, 1.8 μm particle size; Waters), was operated at a temperature of 40°C. The mobile phases consisted of 0.1% formic acid in water (Solvent A) and 0.1% formic acid in acetonitrile (Solvent B). The flow rate of the mobile phase was 400 μl/min, and a 2-μl sample was loaded per injection. The following gradient profile was applied: After 1 min of isocratic run at 99% A, a linear 12-min gradient was applied to 65% A immediately followed by a 1.5-min gradient to 30% A, before a 1-min gradient to 1% A. There then followed a 1.5-min isocratic period at 1% A before switching back to 99% A to re-equilibrate the column for 2.5 min, before the next sample could be injected.

The UPLC was connected to an Exactive Orbitrap (Thermo Fisher Scientific, Bremen, Germany) via a heated electrospray source (Thermo Fisher Scientific). The spectra were recorded using full scan mode, covering a mass range from *m/z* 100 to 1500. The resolution was set to 25 000 and the maximum scan time was set to 250 ms. The sheath gas was set to a value of 60 while the auxiliary gas was set to 35. The transfer capillary temperature was set to 150°C while the heater temperature was adjusted to 300°C. The spray voltage was fixed at 3 kV, with a capillary voltage and a skimmer voltage of 25 and 15 V, respectively. MS spectra were recorded from minute 0 to 19 of the UPLC gradient. Molecular masses, retention times, and associated peak intensities were extracted from the raw files using the RefinerMS Software (Version 5.3; GeneData, Basel, Switzerland) and Xcalibur software (ThermoQuest). Metabolite identification and annotation were performed using standard compounds and a tomato metabolomics database (Moco et al., 2006; Iijima et al., 2008; Tohge and Fernie, 2010). Data are reported in a manner compliant with the standards suggested by Fernie et al. (2011). HCA maps were calculated using MeV software (http://www.tm4.org/) with Pearson's correlation.

Peak identification and annotations were performed using a combinatorial phytochemistry approach. Here, several different strategies were undertaken: (1) hypothetical biosynthetic pathways were postulated based on the chemical structure of known polyphenolic metabolites and peaks were predicted by calculating mass shifts of common metabolite modifications (Morreel et al., 2014; Naake and Fernie, 2019); (2) tomato peel

extracts and the tomato metabolite databases KomicMarket and MotoDB were used for cross-referencing; (3) cross-species comparison with peak annotations from tobacco extracts were carried out (Ruprecht et al., 2016); (4) similarly, cross-species and cross-tissue analysis was carried out by evaluation of peak annotations of floral extracts from Col-0 and *Arabidopsis* flavonoid mutants (Tohge et al., 2016); and finally, (5) the full analytical capacity of the machine was exploited by acquiring four different mass spectra, i.e., in positive/negative ion detection mode and with/without in-source fragmentation, which were used for peak annotation and identification (Supplemental Table 4).

Derivatization and Analysis of Primary Metabolites Using GC–MS

Metabolite extraction for GC–MS was performed using a method modified from that described by Roessner-Tunali et al. (2003). The extraction, derivatization, internal standard addition, and sample injection were exactly as described previously (Lisec et al., 2006). Both chromatograms and mass spectra were evaluated using either TAGFINDER (Luedemann et al., 2012) or Xcalibur software (ThermoQuest), and the resulting data were prepared and presented as described by Roessner et al. (2001). Data are reported in a manner compliant with the standards suggested by Fernie et al. (2011).

Genome-wide Assignment of Orthologous Genes of Specialized Metabolism in *S. lycopersicum*

The enzymatic steps putatively involved in tomato-specialized metabolism were taken from predicted pathways based on the detected metabolites. Genome-wide assignment of tomato orthologous genes was performed via SOLBLAST (http://solgenomics.net/tools/blast/index.pl) using ITAG Release3.2 predicted proteins and using the sequences of experimentally characterized *Arabidopsis* genes as queries. The amino acid sequences of genes absent in *Arabidopsis* such as *p*-coumaroyl-CoA reductase and flavonoid-*C*-glucosyltransferase were obtained from rice and apple. Characterized tomato genes such as chalcone synthase were converted from GenBank ID to ITAG3.2 gene identifier by SOL BLAST search. The returned gene list and amino acid sequences were rechecked by performing a cross-species orthologous cluster search of Plaza (https://bioinformatics.psb.ugent.be/plaza/). All orthologous genes that met the cross-species orthologous cluster search criteria are listed in Table 1.

RNA-Sequencing Analysis of Tomato Tissues and Accessions

RNA-sequencing data were obtained from a published dataset (Koenig et al., 2013; Bolger et al., 2014). Datasets of *S. lycopersicum* tissues (seedling and mature fruits) and tomato accession seedlings (*S. lycopersicum*, *S. pimpinellifolium*, *S. habrochaites*, and *S. pennellii*) and mature fruits (*S. pennellii*) were used. The value of number of reads was normalized by the total number of reads of gene sequences.

Microarray Analysis

Transcriptome analysis was carried out using TOM2. All raw microarray data are available for public download at the Tomato Functional Genomics Database, and processed data can be viewed and queried via the same site (http://ted.bti.cornell.edu). Unigene identifiers were converted to ITAG 2.3 gene identifiers by BLAST search.

Full-Length cDNA of the *SlUGT78-A* (Solyc10g083440) Gene

Full-length cDNAs of the Solyc10g083440 genes from four tomato species (LYCO, PIMP, NEOR, HABR) were cloned and sequenced. Primers used for amplification and sequencing of Solyc10g083440 are 440-f: ATG ACA AGT CCT CAA CTT C and 440-r: TTA AGT AGG CTT GTG ACA T. Both primers were designed using NCBI Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/).

Overexpression of the *SlUGT78-a* Gene in *Arabidopsis*

The SlSolyc10g083440 overexpression construct was created by cloning the full-length cDNA of the Solyc10g083440 gene from mature fruit of *S. lycopersicum* under the control of the CaMV 35S promoter in pK7GW2 (Invitrogen), a binary vector with a Gateway cassette, using the In-Fusion HD cloning kit (Takara). Binary plasmids were transferred to *Agrobacterium tumefaciens* GV3101 (pMP90) and transformed into *Arabidopsis* plants (Col-0) and T-DNA insertion lines (*f3gt*, At5g17050, SALK_049338) (Tohge et al., 2005) according to the floral-dip method. Transgenic plants were grown in the presence of 50 mg/l kanamycin sulfate for pK7GW2 selection, and T4 progenies were used for analysis. The primers used for cloning are SlSolyc10g083440_FW: GGG GAC AAG TTT GTA CAA AAA AGC AGG CTC CAC CAT GAC AAG TCC TCA ACT TCA TAT TG and SlSolyc10g083440-REV: GGG GAC CAC TTT GTA CAA GAA AGC TGG GTC TTA AGT AGG CTT GTG ACA TTT AAT TAG C.

Assay of Recombinant F3GlcT Protein (SlUGT78-a, Solyc10g083440)

The full-length cDNA of Solyc10g083440 from *S. lycopersicum* was cloned into the pCold1 expression vector to fuse it with a histidine tag. The recombinant protein was expressed in SoluBL21 (Genlantis) competent cells hosting pRARE plasmids by following the manufacturer's protocol, then extracted with 20 mM Tris–HCl lysis buffer (pH 7.5), purified by nickel affinity chromatography (Ni Sepharose 6 Fast Flow, GE Healthcare), and concentrated using an Amicon Ultra 0.5 centrifugal filter device. The purified protein in the final suspension buffer (0.1 M sodium phosphate buffer with 10 mM NaCl [pH 7.4]) was subsequently used for the enzyme activity assay, which comprised 0.1 M Tris–HCl (pH 7.4), 5 mM $MgCl_2$, 0.1 mM quercetin, 0.75 mM UDP-glucose or UDP-galactose, and 1 μg of recombinant protein. The 20-μl reaction mixture was incubated at 37°C for 30 min and the reaction was terminated by the addition of 10 μl of 2 M HCl. The glycosyltransferase activity of the recombinant protein was confirmed by LC–MS analysis.

Virus-Induced Gene-Silencing Experiment

Methods of vector construction and infiltration for VIGS were described previously (Alseekh et al., 2015). The gene fragments of Solyc10g083440 and Solyc10g079350 were amplified using Gateway-compatible primers and recombined into the pDONR207 vector (Invitrogen) in a BP reaction to generate an entry clone. The expression clones pTRV2-Solyc10g083440 and pTRV2-Solyc10g079350 were produced by recombining the Entry vector with the pTRV2-GW destination vector using an LR reaction. *A. tumefaciens* strain GV3101:pMP90 was transformed with the sequenced expression vectors by electroporation. Agroinfiltration was performed using the methods described previously (Alseekh et al., 2015). To infiltrate fruit for VIGS, we used MicroTom tomato. The pTRV1 culture and the pTRV2-Solyc10g083440 or pTRV2-Solyc10g079350 culture were mixed in a 1:1 ratio. Fruits were labeled, and the fruit peduncle was injected with 0.2–0.5 ml of bacterial mixture. Fruits were agroinfiltrated at the breaker stage, and samples were harvested at 2 weeks after injection.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at *Molecular Plant Online*.

FUNDING

T.T and A.R.F. gratefully acknowledge partial support by the Max Planck Society and NAIST (to T.T.) as well as the European Union Projects (TOMGEM, MultiBioPro, and PlantaSyst). Research activity of T.T. was additionally supported by the Alexander von Humboldt Foundation (7000228060 to T.T.) and the JSPS KAKENHI Grant-in-Aid for Scientific Research B (19H03249 to T.T.) and C (19K06723 to M.W.).

AUTHOR CONTRIBUTIONS

A.R.F. directed the project. T.T., F.S. and A.R.F. wrote the manuscript. T.T., R.W., P.F., I.B., M.W., S.A., S.S.J., P.G. and J.C.D. performed experiments with the help of Y.Z., J.L. and M.B. T.T., R.W., P.F., M.W., M.L., B.U. performed data analysis and sequence analysis. All authors discussed and interpreted the results.

ACKNOWLEDGMENTS

We are grateful to Dr. Koh Aoki for providing us previously published tomato peel extracts (Iijima et al., 2007) and Dr. Zhangjun Fei for help with the tomato gene conversions.

Received: April 15, 2019
Revised: February 1, 2020
Accepted: April 11, 2020
Published: April 16, 2020

REFERENCES

Afendi, F.M., Okada, T., Yamazaki, M., Hirai-Morita, A., Nakamura, Y., Nakamura, K., Ikeda, S., Takahashi, H., Altaf-Ul-Amin, M., Darusman, K.,L., et al. (2012). KNApSAcK family databases: integrated metabolite–plant species databases for multifaceted plant research. Plant Cell Physiol. **53**:e1.

Alseekh, S., and Fernie, A.R. (2018). Metabolomics 20 years on: what have we learned and what hurdles remain? Plant J. **94**:933–942.

Alseekh, S., Tong, H., Scossa, F., Brotman, Y., Vigroux, F., Tohge, T., Ofner, I., Zamir, D., Nikoloski, Z., and Fernie, A.R. (2017). Canalization of tomato fruit metabolism. Plant Cell **29**:2753–2765.

Alseekh, S., Tohge, T., Wendenberg, R., Scossa, F., Omranian, N., Li, J., Kleessen, S., Giavalisco, P., Pleban, T., Mueller-Roeber, B., et al. (2015). Identification and mode of inheritance of quantitative trait loci for secondary metabolite abundance in tomato. Plant Cell **27**:485–512.

Angelovici, R., Lipka, A.E., Deason, N., Gonzalez-Jorge, S., Lin, H., Cepela, J., Buell, R., Gore, M.A., and Dellapenna, D. (2013). Genome-wide analysis of branched-chain amino acid levels in *Arabidopsis* seeds. Plant Cell **25**:4827–4843.

Ballester, A.R., Molthoff, J., de Vos, R., Hekkert, B., Orzaez, D., Fernandez-Moreno, J.P., Tripodi, P., Grandillo, S., Martin, C., Heldens, J., et al. (2010). Biochemical and molecular analysis of pink tomatoes: deregulated expression of the gene encoding transcription factor SlMYB12 leads to pink tomato fruit color. Plant Physiol. **152**:71–84.

Bellucci, E., Bitocchi, E., Ferrarini, A., Benazzo, A., Biagetti, E., Klie, S., Minio, A., Rau, D., Rodriguez, M., Panziera, A., et al. (2014). Decreased nucleotide and expression diversity and modified coexpression patterns characterize domestication in the common bean. Plant Cell **26**:1901–1912.

Boden, S.A., Cavanagh, C., Cullis, B.R., Ramm, K., Greenwood, J., Finnegan, E.J., Trevaskis, B., and Swain, S.M. (2015). Ppd-1 is a key regulator of inflorescence architecture and paired spikelet development in wheat. Nat. Plants **1**:14016.

Bolger, A., Scossa, F., Bolger, M.E., Lanz, C., Maumus, F., Tohge, T., Quesneville, H., Alseekh, S., Sorensen, I., Lichtenstein, G., et al. (2014). The genome of the stress-tolerant wild tomato species *Solanum pennellii*. Nat. Genet. **46**:1034–1038.

Bongue-Bartelsman, M., O'Neill, S.D., Tong, Y., and Yoder, J.I. (1994). Characterization of the gene encoding dihydroflavonol 4-reductase in tomato. Gene **138**:153–157.

Brazier-Hicks, M., Evans, K.M., Gershater, M.C., Puschmann, H., Steel, P.G., and Edwards, R. (2009). The C-glycosylation of flavonoids in cereals. J. Bio Chem. **284**:17926–17934.

Brockington, S.F., Yang, Y., Gandia-Herrero, F., Covshoff, S., Hibberd, J.M., Sage, R.F., Wong, G.K., Moore, M.J., and Smith, S.A. (2015). Lineage-specific gene radiations underlie the evolution of novel betalain pigmentation in Caryophyllales. New Phytol. **207**:1170–1180.

Butelli, E., Titta, L., Giorgio, M., Mock, H.P., Matros, A., Peterek, S., Schijlen, E.G., Hall, R.D., Bovy, A.G., Luo, J., et al. (2008). Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. Nat. Biotechnol. **26**:1301–1308.

Carrari, F., Baxter, C., Usadel, B., Urbanczyk-Wochniak, E., Zanor, M.I., Nunes-Nesi, A., Nikiforova, V., Centero, D., Ratzka, A., Pauly, M., et al. (2006). Integrated analysis of metabolite and transcript levels reveals the metabolic shifts that underlie tomato fruit development and highlight regulatory aspects of metabolic network behavior. Plant Physiol. **142**:1380–1396.

Chakrabarti, M., Zhang, N., Sauvage, C., Muños, S., Blanca, J., Cañizares, J., Diez, M.J., Schneider, R., Mazourek, M., McClead, J., et al. (2013). A cytochrome P450 regulates a domestication trait in cultivated tomato. Proc. Natl. Acad. Sci. U S A **110**:17125–17130.

Chan, E.K., Rowe, H.C., Corwin, J.A., Joseph, B., and Kliebenstein, D.J. (2011). Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in *Arabidopsis thaliana*. PLoS Biol. **9**:e1001125.

Charaux, C. (1924). Presence of Rutin in certain plants—preparation and identification of this glucoside and of its decomposition products. Bull. Soc. Chim. Biol. **6**:641–647.

Chen, W., Wang, W., Peng, M., Gong, L., Gao, Y., Wan, J., Wang, S., Shi, L., Zhou, B., Li, Z., et al. (2016). Comparative and parallel genome-wide association studies for metabolic and agronomic traits in cereals. Nat. Commun. **7**:12767.

Child, A. (1990). A synopsis of Solanum subgenus Potatoe. Feddes Repert. **101**:209–235.

Clé, C., Hill, L.M., Niggeweg, R., Martin, C.R., Guisez, Y., Prinsen, E., and Jansen, M.A. (2008). Modulation of chlorogenic acid biosynthesis in *Solanum lycopersicum*; consequences for phenolic accumulation and UV-tolerance. Phytochem **69**:2149–2156.

Covey, P.A., Kondo, K., Welch, L., Frank, E., Sianta, S., Kumar, A., Nunez, R., Lopez-Casado, G., van der Knaap, E., Rose, J.K.C., McClure, B.A., and Bedinger, P.A. (2011). Multiple features that distinguish unilateral incongruity and self-incompatibility in the tomato clade. Plant J. **64**:367–378.

Dal Cin, V., Tieman, D.M., Tohge, T., McQuinn, R., de Vos, R.C., Osorio, S., Schmelz, E.A., Taylor, M.G., Smits-Kroon, M.T., Schuurink, R.C., et al. (2011). Identification of genes in the phenylalanine metabolic pathway by ectopic expression of a MYB transcription factor in tomato fruit. Plant Cell **23**:2738–2753.

Dare, A.P., Tomes, S., Jones, M., McGhie, T.K., Stevenson, D.E., Johnson, R.A., Greenwood, D.R., and Hellens, R.P. (2013). Phenotypic changes associated with RNA interference silencing of chalcone synthase in apple (*Malus* × *domestica*). Plant J. **74**:398–410.

Davies, K.M., Albert, N.W., Zhou, Y., and Schwinn, K.E. (2018). Functions of flavonoid and betalain pigments in abiotic stress tolerance in plants. Annu. Plant Rev. Online **1**. https://doi.org/10.1002/9781119312994.apr0604.

Do, P.T., Prudent, M., Sulpice, R., Causse, M., and Fernie, A.R. (2010). The influence of fruit load on the tomato pericarp metabolome in a *Solanum chmielewskii* introgression line population. Plant Physiol. **154**:1128–1142.

Falara, V., Akhtar, T.A., Nguyen, T.T., Spyropoulou, E.A., Bleeker, P.M., Schauvinhold, I., Matsuba, Y., Bonini, M.E., Schilmiller,

A.L., Last, R.L., et al. (2011). The tomato terpene synthase gene family. Plant Physiol. **157**:770–789.

Fan, P., Miller, A.M., Liu, X., Jones, A.D., and Last, R.L. (2017). Evolution of a flipped pathway creates metabolic innovation in tomato trichomes through BAHD enzyme promiscuity. Nat. Commun. **8**:2080.

Fan, P., Miller, A.M., Schilmiller, A.L., Liu, X., Ofner, I., Jones, A.D., Zamir, D., and Last, R.L. (2016). In vitro reconstruction and analysis of evolutionary variation of the tomato acylsucrose metabolic network. Proc. Natl. Acad. Sci. U S A **113**:E239–E248.

Fernie, A.R., and Tohge, T. (2017). The genetics of plant metabolism. Annu. Rev. Genet. **51**:287–310.

Fernie, A.R., and Stitt, M. (2012). On the discordance of metabolomics with proteomics and transcriptomics: coping with increasing complexity in logic, chemistry, and network interactions scientific correspondence. Plant Physiol. **158**:1139–1145.

Fernie, A.R., and Gutierrez-Marcos, J. (2019). From genome to phenome: genome-wide association studies and other approaches that bridge the genotype to phenotype gap. Plant J. **97**:5–7.

Fernie, A.R., Tadmor, Y., and Zamir, D. (2006). Natural genetic variation for improving crop quality. Curr. Opin. Plant Biol. **9**:196–202.

Fernie, A.R., Aharoni, A., Willmitzer, L., Stitt, M., Tohge, T., Kopka, J., Carroll, A.J., Saito, K., Fraser, P.D., and DeLuca, V. (2011). Recommendations for reporting metabolite data. Plant Cell **23**:2477–2482.

Giavalisco, P., Kohl, K., Hummel, J., Seiwert, B., and Willmitzer, L. (2009). ^{13}C isotope-labeled metabolomes allowing for improved compound annotation and relative quantification in liquid chromatography-mass spectrometry-based metabolomic research. Anal. Chem. **81**:6546–6551.

Gomez Roldan, M.V., Outchkourov, N., van Houwelingen, A., Lammers, M., Romero de la Fuente, I., Ziklo, N., Aharoni, A., Hall, R.D., and Beekwilder, J. (2014). An O-methyltransferase modifies accumulation of methylated anthocyanins in seedlings of tomato. Plant J. **80**:695–708.

Grumet, R., Fobes, J.F., and Herner, R.C. (1981). Ripening behavior of wild tomato species. Plant Physiol. **68**:1428–1432.

Haak, D.C., Kostyun, J.L., and Moyle, L.C. (2014). Merging ecology and genomics to dissect diversity in wild tomatoes and their relatives. In Advances in Experimental Medicine and Biology, A.-H.N.E.. and C. Landry, eds. (Dordrecht: Springer).

Iijima, Y., Watanabe, B., Sasaki, R., Takenaka, M., Ono, H., Sakurai, N., Umemoto, N., Suzuki, H., Shibata, D., and Aoki, K. (2013). Steroidal glycoalkaloid profiling and structures of glycoalkaloids in wild tomato fruit. Phytochem **95**:145–157.

Iijima, Y., Nakamura, Y., Ogata, Y., Tanaka, K., Sakurai, N., Suda, K., Suzuki, T., Suzuki, H., Okazaki, K., Kitayama, M., et al. (2008). Metabolite annotations based on the integration of mass spectral information. Plant J. **54**:949–962.

Ishihara, H., Tohge, T., Viehover, P., Fernie, A.R., Weisshaar, B., and Stracke, R. (2016). Natural variation in flavonol accumulation in *Arabidopsis* is determined by the flavonol glucosyltransferase BGLU6. J. Exp. Bot. **67**:1505–1517.

Itkin, M., Heinig, U., Tzfadia, O., Bhide, A.J., Shinde, B., Cardenas, P.D., Bocobza, S.E., Unger, T., Malitsky, S., Finkers, R., et al. (2013). Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. Science **341**:175–179.

Ito, T., Fujimoto, S., Suito, F., Shimosaka, M., and Taguchi, G. (2017). C-Glycosyltransferases catalyzing the formation of di-C-glucosyl flavonoids in citrus plants. Plant J. **91**:187–198.

Jugdé, H., Nguy, D., Moller, I., Cooney, J.M., and Atkinson, R.G. (2008). Isolation and characterization of a novel glycosyltransferase that converts phloretin to phlorizin, a potent antioxidant in apple. FEBS J. **275**:3804–3814.

Keurentjes, J.J., Fu, J., de Vos, C.H., Lommen, A., Hall, R.D., Bino, R.J., van der Plas, L.H., Jansen, R.C., Vreugdenhil, D., and Koornneef, M. (2006). The genetics of plant metabolism. Nat. Genet. **38**:842–849.

Klee, H.J., and Giovannoni, J.J. (2011). Genetics and control of tomato fruit ripening and quality attributes. Annu. Rev. Genet. **45**:41–59.

Kliebenstein, D. (2009). Advancing genetic theory and application by metabolic quantitative trait loci analysis. Plant Cell **21**:1637–1646.

Koenig, D., Jimenez-Gomez, J.M., Kimura, S., Fulop, D., Chitwood, D.H., Headland, L.R., Kumar, R., Covington, M.F., Devisetty, U.K., Tat, A.V., et al. (2013). Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. Proc. Natl. Acad. Sci. U S A **110**:E2655–E2662.

Li, T., Yang, X., Yu, Y., Si, X., Zhai, X., Zhang, H., Dong, W., Gao, C., and Xu, C. (2018a). Domestication of wild tomato is accelerated by genome editing. Nat. Biotechnol. https://doi.org/10.1038/nbt.4273.

Li, Y., Wang, H., Zhang, Y., and Martin, C. (2018b). Can the world's favorite fruit, tomato, provide an effective biosynthetic chassis for high-value metabolites? Plant Cell Rep. **37**:1443–1450.

Lin, T., Zhu, G., Zhang, J., Xu, X., Yu, Q., Zheng, Z., Zhang, Z., Lun, Y., Li, S., Wang, X., et al. (2014). Genomic analyses provide insights into the history of tomato breeding. Nat. Genet. **46**:1220–1226.

Linnaeus, C. (1753). Species plantarum, exhibentes plantas rite cognitas, ad genera relatas, cum differentiis specificis, nominibus trivialibus, synonymis selectis, locis natalibus, secundum systema sexuale digestas (Stockholm: L. Salvius).

Lisec, J., Schauer, N., Kopka, J., Willmitzer, L., and Fernie, A.R. (2006). Gas chromatography mass spectrometry-based metabolite profiling in plants. Nat. Protoc. **1**:387–396.

Luedemann, A., von Malotky, L., Erban, A., and Kopka, J. (2012). TagFinder: preprocessing software for the fingerprinting and the profiling of gas chromatography-mass spectrometry based metabolome analyses. Methods Mol. Biol. **860**:255–286.

Luo, J., Butelli, E., Hill, L., Parr, A., Niggeweg, R., Bailey, P., Weisshaar, B., and Martin, C. (2008). AtMYB12 regulates caffeoyl quinic acid and flavonol synthesis in tomato: expression in fruit results in very high levels of both types of polyphenol. Plant J. **56**:316–326.

Martin-Tanguy, J., Cabanne, F., Perdrizet, E., and Martin, C. (1978). The distribution of hydroxycinnamic acid amides in flowering plants. Phytochem **17**:1927–1928.

Martin, C., Butelli, E., Petroni, K., and Tonelli, C. (2011). How can research on plants contribute to promoting human health? Plant Cell **23**:1685–1699.

McCouch, S. (2004). Diversifying selection in plant breeding. PLoS Biol. **2**:e347.

McCouch, S., Baute, G.J., Bradeen, J., Bramel, P., Bretting, P.K., Buckler, E., Burke, J.M., Charest, D., Cloutier, S., Cole, G., et al. (2013). Agriculture: feeding the future. Nature **499**:23–24.

Mintz-Oron, S., Mandel, T., Rogachev, I., Feldberg, L., Lotan, O., Yativ, M., Wang, Z., Jetter, R., Venger, I., Adato, A., et al. (2008). Gene expression and metabolism in tomato fruit surface tissues. Plant Physiol. **147**:823–851.

Mitchell-Olds, T., Feder, M., and Wray, G. (2008). Evolutionary and ecological functional genomics. Heredity **100**:101–102.

Moco, S., Capanoglu, E., Tikunov, Y., Bino, R.J., Boyacioglu, D., Hall, R.D., Vervoort, J., and De Vos, R.C. (2007). Tissue specialization at

the metabolite level is perceived during the development of tomato fruit. J. Exp. Bot. **58**:4131–4146.

Moco, S., Bino, R.J., Vorst, O., Verhoeven, H.A., de Groot, J., van Beek, T.A., Vervoort, J., and de Vos, C.H.R. (2006). A liquid chromatography-mass spectrometry-based metabolome database for tomato. Plant Physiol. **141**:563–568.

Moghe, G.D., Leong, B.J., Hurney, S.M., Daniel Jones, A., and Last, R.L. (2017). Evolutionary routes to biochemical innovation revealed by integrative analysis of a plant-defense related specialized metabolic pathway. eLife **6**. https://doi.org/10.7554/eLife.28468.

Morreel, K., Saeys, Y., Dima, O., Lu, F., Van de Peer, Y., Vanholme, R., Ralph, J., Vanholme, B., and Boerjan, W. (2014). Systematic structural characterization of metabolites in *Arabidopsis* via candidate substrate-product pair networks. Plant Cell **26**:929–945.

Müller, N.A., Zhang, L., Koornneef, M., and Jiménez-Gómez, J.M. (2018). Mutations in EID1 and LNK2 caused light-conditional clock deceleration during tomato domestication. Proc. Natl. Acad. Sci. U S A **115**:7135–7140.

Müller, N.A., Wijnen, C.L., Srinivasan, A., Ryngajllo, M., Ofner, I., Lin, T., Ranjan, A., West, D., Maloof, J.N., Sinha, N.R., et al. (2016). Domestication selected for deceleration of the circadian clock in cultivated tomato. Nat. Genet. **48**:89–93.

Naake, T., and Fernie, A.R. (2019). MetNet: metabolite network prediction from high-resolution mass spectrometry data in R aiding metabolite annotation. Anal. Chem. **91**:1768–1772.

Nakabayashi, R., Yonekura-Sakakibara, K., Urano, K., Suzuki, M., Yamada, Y., Nishizawa, T., Matsuda, F., Kojima, M., Sakakibara, H., Shinozaki, K., et al. (2014). Enhancement of oxidative and drought tolerance in *Arabidopsis* by overaccumulation of antioxidant flavonoids. Plant J. **77**:367–379.

Nashilevitz, S., Melamed-Bessudo, C., Izkovich, Y., Rogachev, I., Osorio, S., Itkin, M., Adato, A., Pankratov, I., Hirschberg, J., Fernie, A.R., et al. (2010). An orange ripening mutant links plastid NAD(P)H dehydrogenase complex activity to central and specialized metabolism during tomato fruit maturation. Plant Cell **22**:1977–1997.

Niggeweg, R., Michael, A.J., and Martin, C. (2004). Engineering plants with increased levels of the antioxidant chlorogenic acid. Nat. Biotechnol. **22**:746–754.

O'Neill, S.D., Tong, Y., Spörlein, B., Forkmann, G., and Yoder, J.I. (1990). Molecular genetic analysis of chalcone synthase in *Lycopersicon esculentum* and an anthocyanin-deficient mutant. Mol. Gen. Genet. **224**:279–288.

Ogata, J., Itoh, Y., Ishida, M., Yoshida, H., and Ozeki, Y. (2004). Cloning and heterologous expression of cDNAs encoding flavonoid glucosyltransferases from Dianthus caryophyllus. Plant Biotechnol. **21**:367–375.

Okazaki, Y., and Saito, K. (2016). Integrated metabolomics and phytochemical genomics approaches for studies on rice. Gigascience **5**:11.

Ono, E., Fukuchi-Mizutani, M., Nakamura, N., Fukui, Y., Yonekura-Sakakibara, K., Yamaguchi, M., Nakayama, T., Tanaka, T., Kusumi, T., and Tanaka, Y. (2006). Yellow flowers generated by expression of the aurone biosynthetic pathway. Proc. Natl. Acad. Sci. U S A **103**:11075–11080.

Orzaez, D., Medina, A., Torre, S., Fernandez-Moreno, J.P., Rambla, J.L., Fernandez-Del-Carmen, A., Butelli, E., Martin, C., and Granell, A. (2009). A visual reporter system for virus-induced gene silencing in tomato fruit based on anthocyanin accumulation. Plant Physiol. **150**:1122–1134.

Peralta, E., and Spooner, D. (2000). Classification of wild tomatoes: a review. Kurtziana **28**:45–54.

Peralta, I.E., Spooner, D.M., and Knapp, S. (2008). Taxonomy of wild tomatoes and their relatives (Solanum sect Lycopersicoides, sect. Juglandifolia, sect. Lycopersicon; Solanaceae). Syst. Bot. Monogr. **84**:1–186.

Perez-Fons, L., Wells, T., Corol, D.I., Ward, J.L., Gerrish, C., Beale, M.H., Seymour, G.B., Bramley, P.M., and Fraser, P.D. (2014). A genome-wide metabolomic resource for tomato fruit from *Solanum pennellii*. Sci. Rep. **4**:3859.

Perez de Souza, L., Scossa, F., Proost, S., Bitocchi, E., Papa, R., Tohge, T., and Fernie, A.R. (2019). Multi-tissue integration of transcriptomic and specialized metabolite profiling provides tools for assessing the common bean (*Phaseolus vulgaris*) metabolome. Plant J. **97**:1132–1153.

Platt, A., Horton, M., Huang, Y.S., Li, Y., Anastasio, A.E., Mulyati, N.W., Agren, J., Bossdorf, O., Byers, D., Donohue, K., et al. (2010). The scale of population structure in *Arabidopsis thaliana*. Plos Genet. **6**:e1000843.

Riedelsheimer, C., Czedik-Eysenberg, A., Grieder, C., Lisec, J., Technow, F., Sulpice, R., Altmann, T., Stitt, M., Willmitzer, L., and Melchinger, A.E. (2012). Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nat. Genet. **44**:217–220.

Roessner-Tunali, U., Hegemann, B., Lytovchenko, A., Carrari, F., Bruedigam, C., Granot, D., and Fernie, A.R. (2003). Metabolic profiling of transgenic tomato plants overexpressing hexokinase reveals that the influence of hexose phosphorylation diminishes during fruit development. Plant Physiol. **133**:84–99.

Roessner, U., Luedemann, A., Brust, D., Fiehn, O., Linke, T., Willmitzer, L., and Fernie, A. (2001). Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. Plant Cell **13**:11–29.

Rohrmann, J., Tohge, T., Alba, R., Osorio, S., Caldana, C., McQuinn, R., Arvidsson, S., van der Merwe, M.J., Riano-Pachon, D.M., Mueller-Roeber, B., et al. (2011). Combined transcription factor profiling, microarray analysis and metabolite profiling reveals the transcriptional control of metabolic shifts occurring during tomato fruit development. Plant J. **68**:999–1013.

Routaboul, J.M., Kerhoas, L., Debeaujon, I., Pourcel, L., Caboche, M., Einhorn, J., and Lepiniec, L. (2006). Flavonoid diversity and biosynthesis in seed of *Arabidopsis thaliana*. Planta **224**:96–107.

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.

Ruprecht, C., Mendrinna, A., Tohge, T., Sampathkumar, A., Klie, S., Fernie, A.R., Nikoloski, Z., Persson, S., and Mutwil, M. (2016). FamNet: a framework to identify multiplied modules driving pathway expansion in plants. Plant Physiol. **170**:1878–1894.

Saito, K., Yonekura-Sakakibara, K., Nakabayashi, R., Higashi, Y., Yamazaki, M., Tohge, T., and Fernie, A.R. (2013). The flavonoid biosynthetic pathway in *Arabidopsis*: structural andgenetic diversity. Plant Physiol. Biochem. **72**:21–34.

Scarano, A., Butelli, E., De Santis, S., Cavalcanti, E., Hill, L., De Angelis, M., Giovinazzo, G., Chieppa, M., Martin, C., and Santino, A. (2017). Combined dietary anthocyanins, flavonols, and stilbenoids alleviate inflammatory bowel disease symptoms in mice. Front. Nutr. **4**:75.

Schauer, N., and Fernie, A.R. (2006). Plant metabolomics: towards biological function and mechanism. Trends Plant Sci. **11**:508–516.

Schauer, N., Zamir, D., and Fernie, A.R. (2005). Metabolic profiling of leaves and fruit of wild species tomato: a survey of the *Solanum lycopersicum* complex. J. Exp. Bot. **56**:297–307.

Schauer, N., Semel, Y., Balbo, I., Steinfath, M., Repsilber, D., Selbig, J., Pleban, T., Zamir, D., and Fernie, A.R. (2008). Mode of

inheritance of primary metabolic traits in tomato. Plant Cell **20**:509–523.

Schauer, N., Semel, Y., Roessner, U., Gur, A., Balbo, I., Carrari, F., Pleban, T., Perez-Melis, A., Bruedigam, C., Kopka, J., et al. (2006). Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. Nat. Biotechnol. **24**:447–454.

Schilmiller, A., Shi, F., Kim, J., Charbonneau, A.L., Holmes, D., Daniel Jones, A., and Last, R.L. (2010). Mass spectrometry screening reveals widespread diversity in trichome specialized metabolites of tomato chromosomal substitution lines. Plant J. **62**:391–403.

Schmidt, A., Li, C., Jones, A.D., and Pichersky, E. (2012). Characterization of a flavonol 3-O-methyltransferase in the trichomes of the wild tomato species *Solanum habrochaites*. Planta **236**:839–849.

Schmidt, A., Li, C., Shi, F., Jones, A.D., and Pichersky, E. (2011). Polymethylated myricetin in trichomes of the wild tomato species *Solanum habrochaites* and characterization of trichome-specific 3′/5′- and 7/4′-myricetin O-methyltransferases. Plant Physiol. **155**:1999–2009.

Schneeberger, K., and Weigel, D. (2011). Fast-forward genetics enabled by new sequencing technologies. Trends Plant Sci. **16**:282–288.

Schreiber, G., Reuveni, M., Evenor, D., Oren-Shamir, M., Ovadia, R., Sapir-Mir, M., Bootbool-Man, A., Nahon, S., Shlomo, H., Chen, L., et al. (2012). ANTHOCYANIN1 from Solanum chilense is more efficient in accumulating anthocyanin metabolites than its Solanum lycopersicum counterpart in association with the ANTHOCYANIN FRUIT phenotype of tomato. Theor. Appl. Genet. **124**:295–307.

Schulz, E., Tohge, T., Zuther, E., Fernie, A.R., and Hincha, D.K. (2016). Flavonoids are determinants of freezing tolerance and cold acclimation in *Arabidopsis thaliana*. Sci. Rep. **6**:34027.

Schwahn, K., de Souza, L.P., Fernie, A.R., and Tohge, T. (2014). Metabolomics-assisted refinement of the pathways of steroidal glycoalkaloid biosynthesis in the tomato clade. J. Integr. Plant Biol. **56**:864–875.

Shahaf, N., Rogachev, I., Heinig, U., Meir, S., Malitsky, S., Battat, M., Wyner, H., Zheng, S., Wehrens, R., and Aharoni, A. (2016). The WEIZMASS spectral library for high-confidence metabolite identification. Nat. Commun. **7**:12423.

Shikazono, N., Yokota, Y., Tanaka, A., Watanabe, H., and Tano, S. (1998). Molecular analysis of carbon ion-induced mutations in *Arabidopsis thaliana*. Genes Genet. Syst. **73**:173–179.

Slimestad, R., Fossen, T., and Verheul, M.J. (2008). The flavonoids of tomatoes. J. Agric. Food Chem. **56**:2436–2441.

Slimestad, R., and Verheul, M. (2011). Properties of chalconaringenin and rutin isolated from cherry tomatoes. J. Agric. Food Chem. **59**:3180–3185.

Steinhauser, M.C., Steinhauser, D., Gibon, Y., Bolger, M., Arrivault, S., Usadel, B., Zamir, D., Fernie, A.R., and Stitt, M. (2011). Identification of enzyme activity quantitative trait loci in a *Solanum lycopersicum* x *Solanum pennellii* introgression line population. Plant Physiol. **157**:998–1014.

Stevens, R., Buret, M., Duffe, P., Garchery, C., Baldet, P., Rothan, C., and Causse, M. (2007). Candidate genes and quantitative trait loci affecting fruit ascorbic acid content in three tomato populations. Plant Physiol. **143**:1943–1953.

Stracke, R., De Vos, R.C., Bartelniewoehner, L., Ishihara, H., Sagasser, M., Martens, S., and Weisshaar, B. (2009). Metabolomic and genetic analyses of flavonol synthesis in *Arabidopsis thaliana* support the in vivo involvement of leucoanthocyanidin dioxygenase. Planta **229**:427–445.

Sulpice, R., Trenkamp, S., Steinfath, M., Usadel, B., Gibon, Y., Witucka-Wall, H., Pyl, E.T., Tschoep, H., Steinhauser, M.C., Guenther, M., 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.

Tanksley, S.D., and McCouch, S.R. (1997). Seed banks and molecular maps: unlocking genetic potential from the wild. Science **277**:1063–1066.

Teutschbein, J., Gross, W., Nimtz, M., Milkowski, C., Hause, B., and Strack, D. (2010). Identification and localization of a lipase-like acyltransferase in phenylpropanoid metabolism of tomato (*Solanum lycopersicum*). J. Biol. Chem. **285**:38374–38381.

Tian, F., Bradbury, P.J., Brown, P.J., Hung, H., Sun, Q., Flint-Garcia, S., Rocheford, T.R., McMullen, M.D., Holland, J.B., and Buckler, E.S. (2011). Genome-wide association study of leaf architecture in the maize nested association mapping population. Nat. Genet. **43**:159–162.

Tieman, D., Zhu, G., Resende, M.F., Jr., Lin, T., Nguyen, C., Bies, D., Rambla, J.L., Beltran, K.S., Taylor, M., Zhang, B., et al. (2017). A chemical genetic roadmap to improved tomato flavor. Science **355**:391–394.

Tieman, D.M., Zeigler, M., Schmelz, E.A., Taylor, M.G., Bliss, P., Kirst, M., and Klee, H.J. (2006). Identification of loci affecting flavour volatile emissions in tomato fruits. J. Exp. Bot. **57**:887–896.

Tikunov, Y., Lommen, A., de Vos, C.H., Verhoeven, H.A., Bino, R.J., Hall, R.D., and Bovy, A.G. (2005). A novel approach for nontargeted data analysis for metabolomics. Large-scale profiling of tomato fruit volatiles. Plant Physiol. **139**:1125–1137.

Tikunov, Y.M., Molthoff, J., de Vos, R.C., Beekwilder, J., van Houwelingen, A., van der Hooft, J.J., Nijenhuis-de Vries, M., Labrie, C.W., Verkerke, W., van de Geest, H., et al. (2013). Non-smoky glycosyltransferase1 prevents the release of smoky aroma from tomato fruit. Plant Cell **25**:3067–3078.

Tohge, T., Yonekura-Sakakibara, K., Niida, R., Watanabe-Takahashi, A., and Saito, K. (2007). Phytochemical genomics in Arabidopsis thaliana: a case study for functional identification of flavonoid biosynthesis genes. Pure Appl Chem **79**:811–823.

Tohge, T., and Fernie, A.R. (2010). Combining genetic diversity, informatics and metabolomics to facilitate annotation of plant gene function. Nat. Protoc. **5**:1210–1227.

Tohge, T., and Fernie, A.R. (2015). Metabolomics-inspired insight into developmental, environmental and genetic aspects of tomato fruit chemical composition and quality. Plant Cell Physiol **56**:1681–1696.

Tohge, T., and Fernie, A.R. (2017). An overview of compounds derived from the shikimate and phenylpropanoid pathways and their medicinal importance. Mini Rev. Med. Chem. **17**:1013–1027.

Tohge, T., de Souza, L.P., and Fernie, A.R. (2014). Genome-enabled plant metabolomics. Analyt. Technol. Biomed. Life Sci. **966**:7–20.

Tohge, T., Watanabe, M., Hoefgen, R., and Fernie, A.R. (2013). The evolution of phenylpropanoid metabolism in the green lineage. Crit. Rev. Biochem. Mol. Biol. **48**:123–152.

Tohge, T., Mettler, T., Arrivault, S., Carroll, A.J., Stitt, M., and Fernie, A.R. (2011). From models to crop species: caveats and solutions for translational metabolomics. Front. Plant Sci. **2**:61.

Tohge, T., Wendenburg, R., Ishihara, H., Nakabayashi, R., Watanabe, M., Sulpice, R., Hoefgen, R., Takayama, H., Saito, K., Stitt, M., et al. (2016). Characterization of a recently evolved flavonol-phenylacyltransferase gene provides signatures of natural light selection in Brassicaceae. Nat. Commun. **7**:12399.

Tohge, T., Zhang, Y., Peterek, S., Matros, A., Rallapalli, G., Tandron, Y.A., Butelli, E., Kallam, K., Hertkorn, N., Mock, H.P., et al. (2015). Ectopic expression of snapdragon transcription factors facilitates the identification of genes encoding enzymes of anthocyanin decoration in tomato. Plant J. **83**:686–704.

Tohge, T., Nishiyama, Y., Hirai, M.Y., Yano, M., Nakajima, J., Awazuhara, M., Inoue, E., Takahashi, H., Goodenowe, D.B., Kitayama, 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.

Tomato Genome Consortium. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature **485**:635–641.

100 Tomato Genome Sequencing Consortium. (2014). Exploring genetic variation in the tomato (*Solanum* section Lycopersicon) clade by whole-genome sequencing. Plant J. **80**:136–148.

Urbanczyk-Wochniak, E., Usadel, B., Thimm, O., Nunes-Nesi, A., Carrari, F., Davy, M., Blasing, O., Kowalczyk, M., Weicht, D., Polinceusz, A., et al. (2006). Conversion of MapMan to allow the analysis of transcript data from Solanaceous species: effects of genetic and environmental alterations in energy metabolism in the leaf. Plant Mol. Biol. **60**:773–792.

Wang, M., Li, W., Fang, C., Xu, F., Liu, Y., Wang, Z., Yang, R., Zhang, M., Liu, S., Lu, S., et al. (2018). Parallel selection on a dormancy gene during domestication of crops from multiple families. Nat. Genet. **50**:1435–1441.

Wen, W., Li, D., Li, X., Gao, Y., Li, W., Li, H., Liu, J., Liu, H., Chen, W., Luo, J., et al. (2014). Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. Nat. Commun. **5**:3438.

Wu, M., and Burrell, R.C. (1958). Flavonoid pigments of the tomato (*Lycopersicum esculentum* Mill.). Arch. Biochem. Biophys. **74**:114–118.

Wu, S., Tohge, T., Cuadros-Inostroza, A., Tong, H., Tenenboim, H., Kooke, R., Meret, M., Keurentjes, J.B., Nikoloski, Z., Fernie, A.R., et al. (2018). Mapping the *Arabidopsis* metabolic landscape by untargeted metabolomics at different environmental conditions. Mol. Plant **11**:118–134.

Ye, J., Wang, X., Hu, T., Zhang, F., Wang, B., Li, C., Yang, T., Li, H., Lu, Y., Giovannoni, J.J., et al. (2017). An InDel in the promoter of Al-ACTIVATED MALATE TRANSPORTER9 selected during tomato domestication determines fruit malate contents and aluminum tolerance. Plant Cell **29**:2249–2268.

Yeats, T.H., Buda, G.J., Wang, Z., Chehanovsky, N., Moyle, L.C., Jetter, R., Schaffer, A.A., and Rose, J.K. (2012a). The fruit cuticles of wild tomato species exhibit architectural and chemical diversity, providing a new model for studying the evolution of cuticle function. Plant J. **69**:655–666.

Yeats, T.H., Martin, L.B., Viart, H.M., Isaacson, T., He, Y., Zhao, L., Matas, A.J., Buda, G.J., Domozych, D.S., Clausen, M.H., et al. (2012b). The identification of cutin synthase: formation of the plant polyester cutin. Nat. Chem. Biol. **8**:609–611.

Yonekura-Sakakibara, K., Tohge, T., Niida, R., and Saito, K. (2007). Identification of a flavonol 7-O-rhamnosyltransferase gene determining flavonoid pattern in *Arabidopsis* by transcriptome coexpression analysis and reverse genetics. J. Biol. Chem. **282**:14932–14941.

Yonekura-Sakakibara, K., Tohge, T., Matsuda, F., Nakabayashi, R., Takayama, H., Niida, R., Watanabe-Takahashi, A., Inoue, E., and Saito, K. (2008). Comprehensive flavonol profiling and transcriptome coexpression analysis leading to decoding gene-metabolite correlations in *Arabidopsis*. Plant Cell **20**:2160–2176.

Zamir, D. (2001). Improving plant breeding with exotic genetic libraries. Nat. Rev. Genet. **2**:983–989.

Zhang, Y., Butelli, E., Alseekh, S., Tohge, T., Rallapalli, G., Luo, J., Kawar, P.G., Hill, L., Santino, A., Fernie, A.R., et al. (2015). Multi-level engineering facilitates the production of phenylpropanoid compounds in tomato. Nat. Commun. **6**:8635.

Zhang, Y., Butelli, E., De Stefano, R., Schoonbeek, H.J., Magusin, A., Pagliarani, C., Wellner, N., Hill, L., Orzaez, D., Granell, A., et al. (2013). Anthocyanins double the shelf life of tomatoes by delaying overripening and reducing susceptibility to gray mold. Curr. Biol. **23**:1094–1110.

Zhu, G., Wang, S., Huang, Z., Zhang, S., Liao, Q., Zhang, C., Lin, T., Qin, M., Peng, M., Yang, C., et al. (2018). Rewiring of the fruit metabolome in tomato breeding. Cell **172**:249–261 e212.

Zsögön, A., Čermák, T., Naves, E.R., Notini, M.M., Edel, K.H., Weinl, S., Freschi, L., Voytas, D.F., Kudla, J., and Peres, L.E.P. (2018). De novo domestication of wild tomato using genome editing. Nat. Biotechnol. https://doi.org/10.1038/nbt.4272.

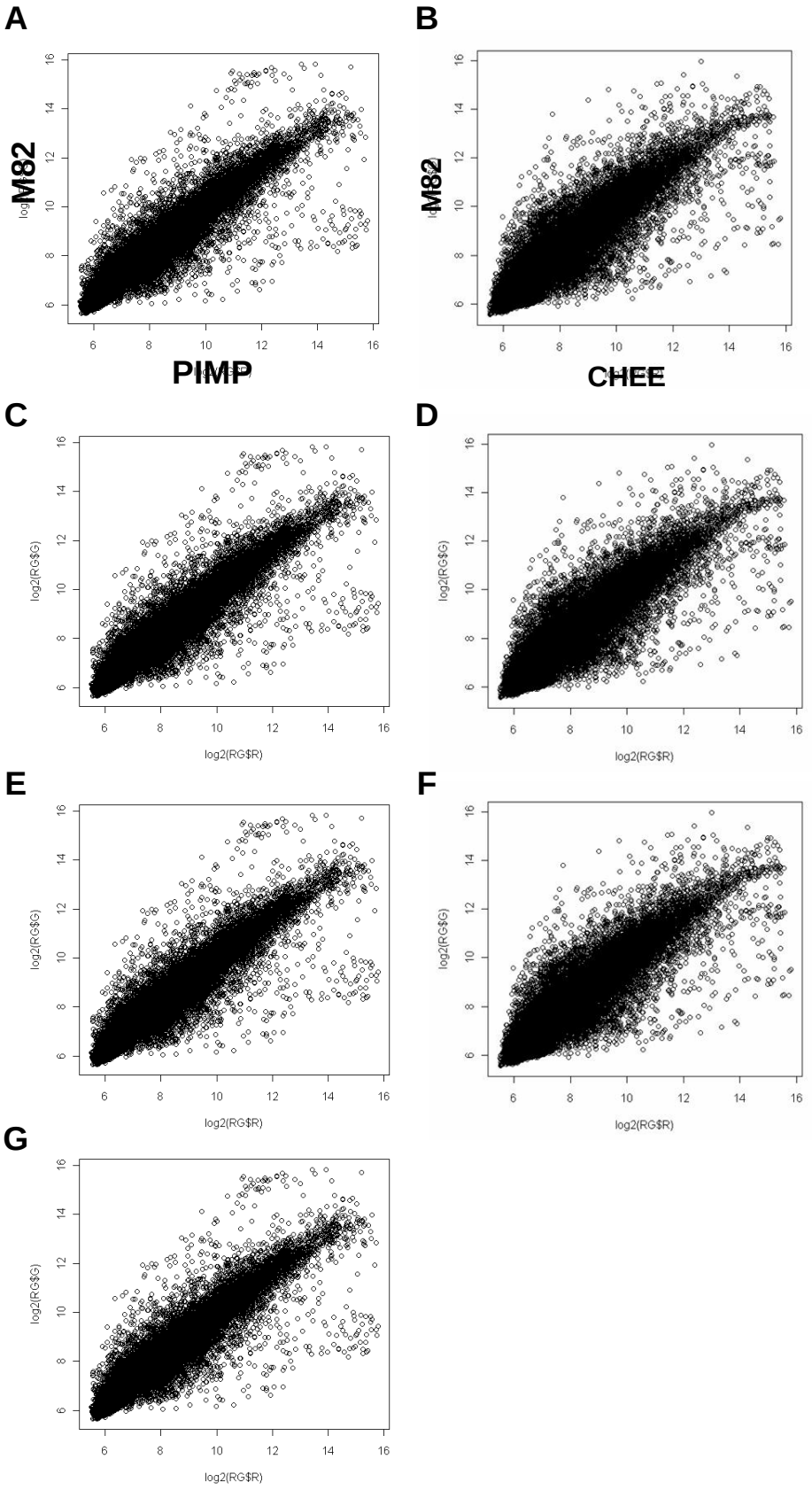
Zuo, J.R., and Li, J.Y. (2014). Molecular genetic dissection of quantitative trait loci regulating rice grain size. Annu. Rev. Genet. **48**:99–118.

Molecular Plant, Volume *13*

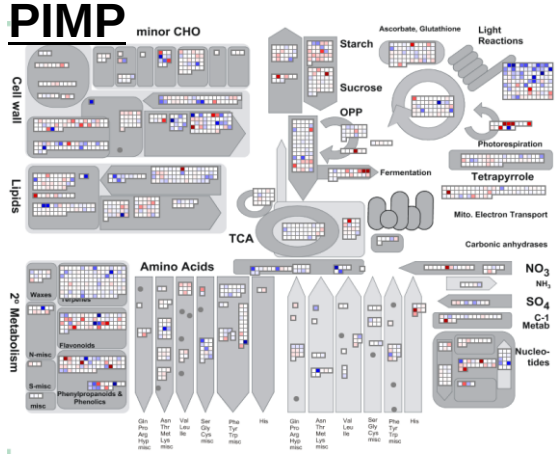
Supplemental Information

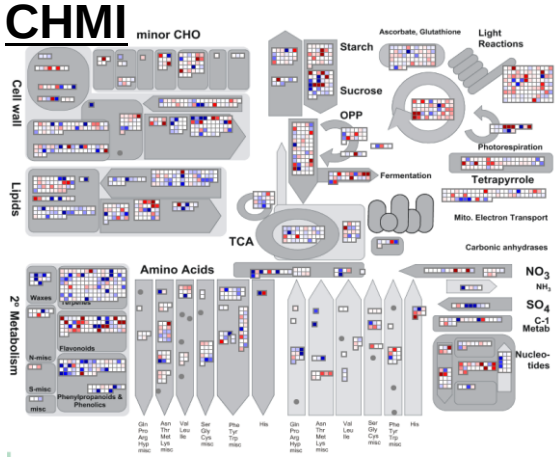
Exploiting Natural Variation in Tomato to Define Pathway Structure and Metabolic Regulation of Fruit Polyphenolics in the Lycopersicum Complex

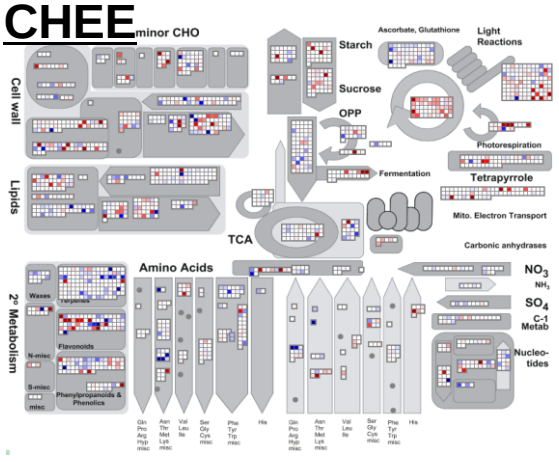
Takayuki Tohge, Federico Scossa, Regina Wendenburg, Pierre Frasse, Ilse Balbo, Mutsumi Watanabe, Saleh Alseekh, Sagar Sudam Jadhav, Jay C. Delfin, Marc Lohse, Patrick Giavalisco, Bjoern Usadel, Youjun Zhang, Jie Luo, Mondher Bouzayen, and Alisdair R. Fernie

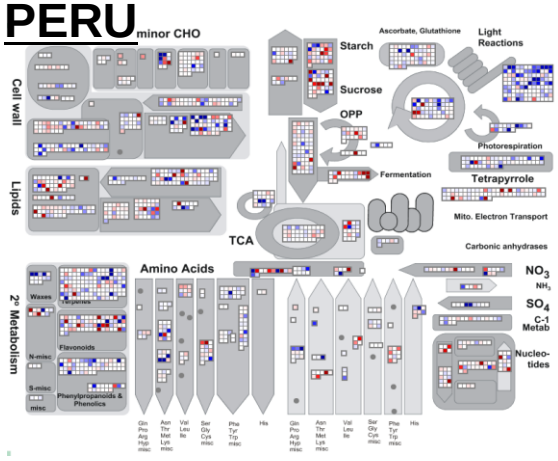


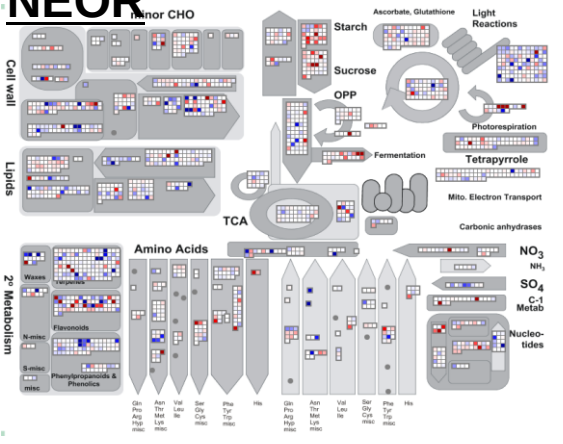
Supplemental Figure 1. Scatter plotting of gene expression of TOM2 microarray for evaluation of cross-hybridization strategy.

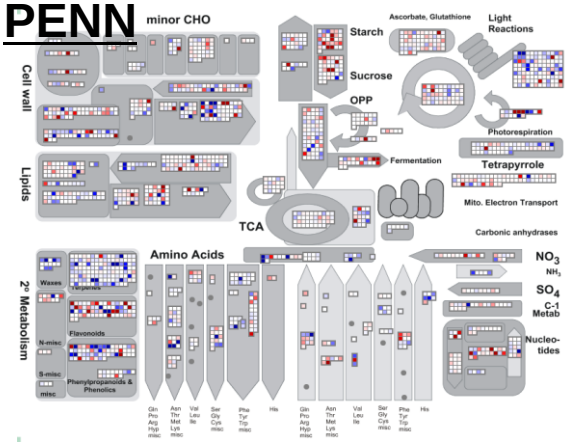


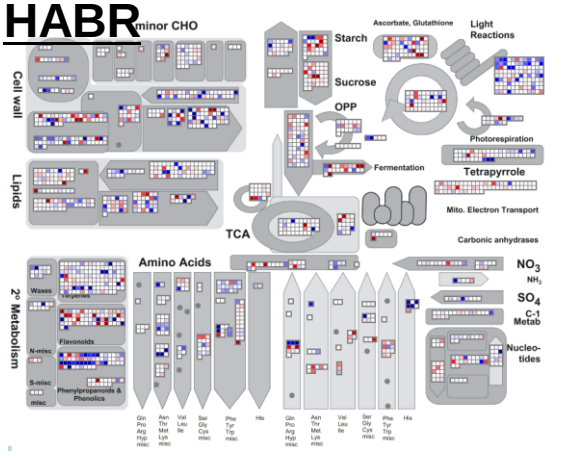








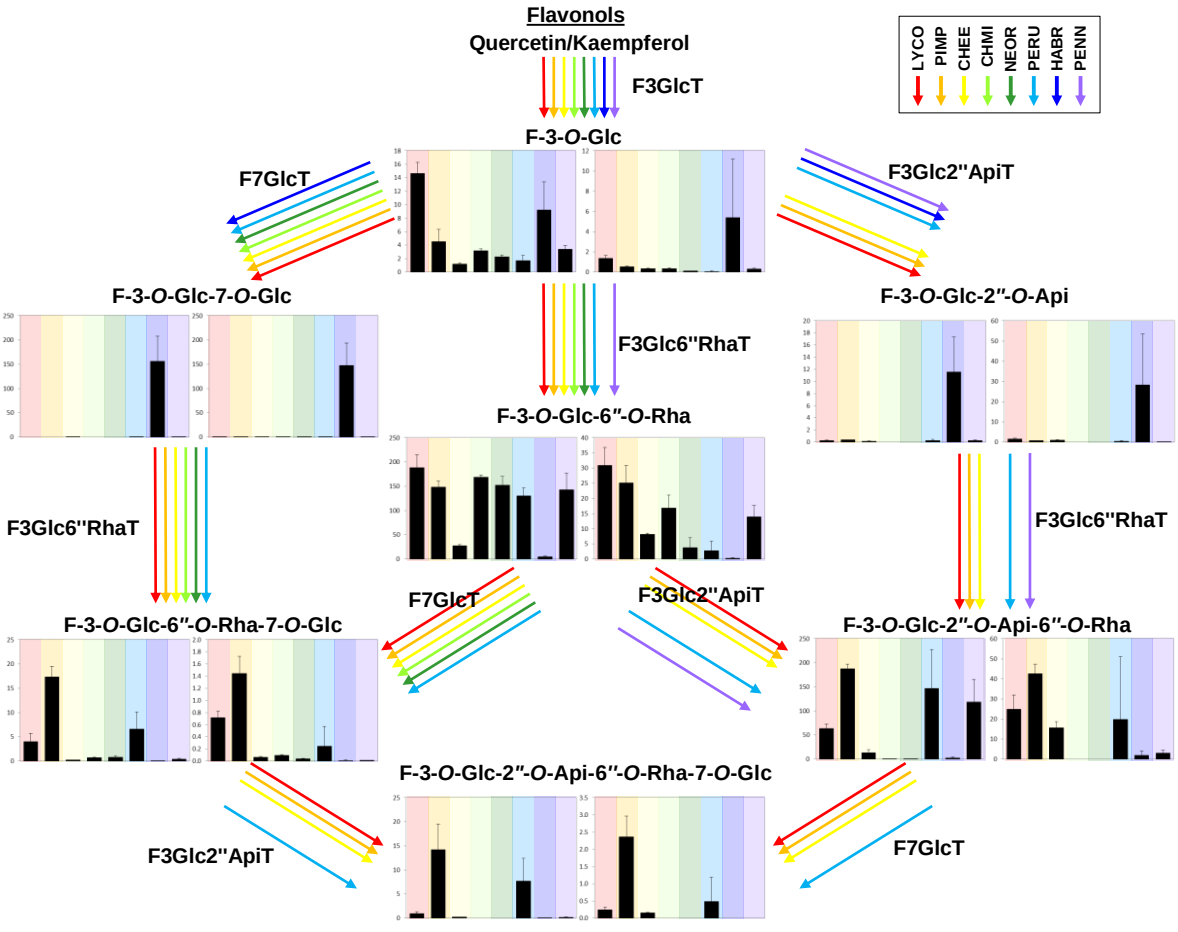




Supplemental Figure 2. Mapman analysis of microarray result.

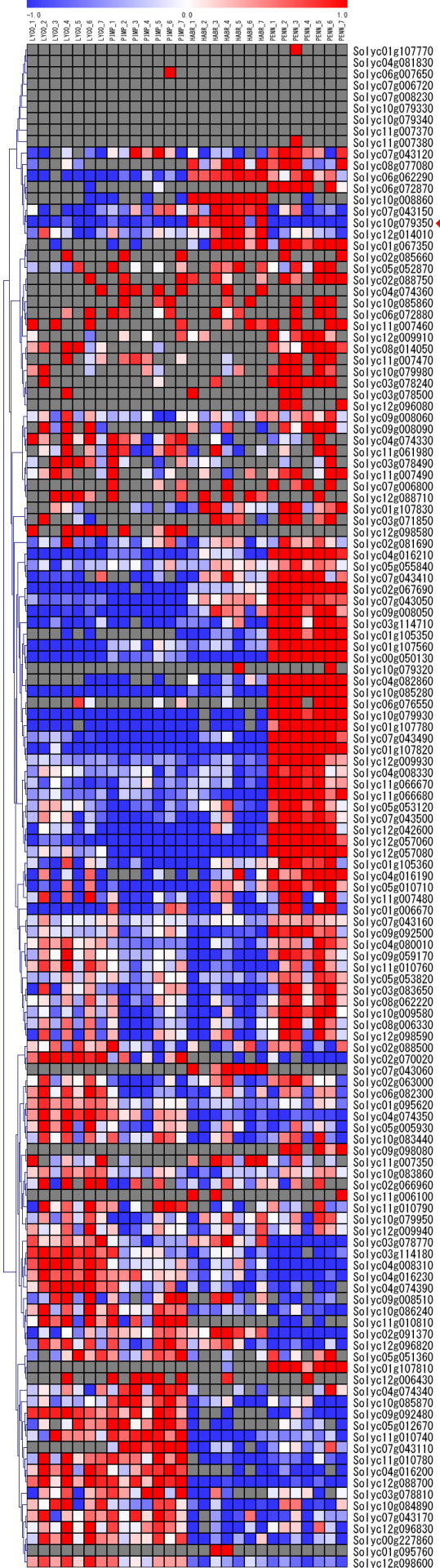
Rt (min)	*m/z*	Compound	CHS	CHI	CDRB	P3'CGT	F3H	FLS	F3'H	OMT	Fd3GT	F2"ApiT	F6"RhaT	F7GT	FAT	NGT
7.13	447.0926	KG	1	1	0	0	1	1	0	0	1	0	0	0	0	0
6.67	579.1348	KGA	1	1	0	0	1	1	0	0	1	1	0	0	0	0
6.14	609.1462	KGG	1	1	0	0	1	1	0	0	1	0	0	1	0	0
6.89	593.1501	KGR	1	1	0	0	1	1	0	0	1	0	1	0	0	0
6.36	725.1929	KGRA	1	1	0	0	1	1	0	0	1	1	1	0	0	0
5.09	755.2031	KGRG	1	1	0	0	1	1	0	0	1	0	1	1	0	0
4.68	887.2452	KGRAG	1	1	0	0	1	1	0	0	1	1	1	1	0	0
7.27	887.2245	KGRACaf	1	1	0	0	1	1	0	0	1	1	1	0	1	0
7.93	901.2408	KGRAFer	1	1	0	0	1	1	0	0	1	1	1	0	1	0
5.77	1049.2780	KGRAGCaf	1	1	0	0	1	1	0	0	1	1	1	1	1	0
6.36	1063.2980	KGRAGFer	1	1	0	0	1	1	0	0	1	1	1	1	1	0
7.90	871.2319	KGRApCou	1	1	0	0	1	1	0	0	1	1	1	0	1	0
7.71	931.2511	KGRASin	1	1	0	0	1	1	0	0	1	1	1	0	1	0
6.56	463.0878	QG	1	1	0	0	1	1	1	0	1	0	0	0	0	0
6.19	595.1315	QGA	1	1	0	0	1	1	1	0	1	1	0	0	0	0
5.68	625.1412	QGG	1	1	0	0	1	1	1	0	1	0	0	1	0	0
6.33	609.1458	QGR	1	1	0	0	1	1	1	0	1	0	1	0	0	0
5.93	741.1879	QGRA	1	1	0	0	1	1	1	0	1	1	1	0	0	0
4.43	903.2402	QGRAG	1	1	0	0	1	1	1	0	1	1	1	1	0	0
4.67	771.1984	QGRG	1	1	0	0	1	1	1	0	1	0	1	1	0	0
6.90	903.2190	QGRACaf	1	1	0	0	1	1	1	0	1	1	1	0	1	0
7.54	917.2356	QGRAFer	1	1	0	0	1	1	1	0	1	1	1	0	1	0
5.49	1065.2740	QGRAGCaf	1	1	0	0	1	1	1	0	1	1	1	1	1	0
5.98	1049.2760	QGRAGpCou	1	1	0	0	1	1	1	0	1	1	1	1	1	0
6.77	933.2288	QGRAHO5Fer	1	1	0	0	1	1	1	0	1	1	1	0	1	0
7.52	887.2274	QGRApCou	1	1	0	0	1	1	1	0	1	1	1	0	1	0
7.33	947.2453	QGRASin	1	1	0	0	1	1	1	0	1	1	1	0	1	0
7.29	477.1030	IsG	1	1	0	0	1	1	1	1	1	0	0	0	0	0
6.79	609.1467	IsGA	1	1	0	0	1	1	1	1	1	1	0	0	0	0
6.22	639.1567	IsGG	1	1	0	0	1	1	1	1	1	0	0	1	0	0
7.02	623.1614	IsGR	1	1	0	0	1	1	1	1	1	0	1	0	0	0
6.44	755.2034	IsGRA	1	1	0	0	1	1	1	1	1	1	1	0	0	0
4.84	917.2581	IsGRAG	1	1	0	0	1	1	1	1	1	1	1	1	0	0
5.22	785.2141	IsGRG	1	1	0	0	1	1	1	1	1	0	1	1	0	0
9.86	271.0611	NariChal	1	0	0	0	0	0	0	0	0	0	0	0	0	0
7.39	433.1136	NariChal-G1	1	0	0	0	0	0	0	0	0	0	0	0	0	1
8.09	433.1136	NariChal-G2	1	0	0	0	0	0	0	0	0	0	0	0	0	1
8.16	433.1133	NariChal-G3	1	0	0	0	0	0	0	0	0	0	0	0	0	1
6.64	597.1833	P35diGlc	1	0	1	1	0	0	0	0	0	0	0	0	0	0

Supplemental Figure 3. Indication of enzymatic genes adn polyphenols for the prediction of gene expression from metabolomic data.



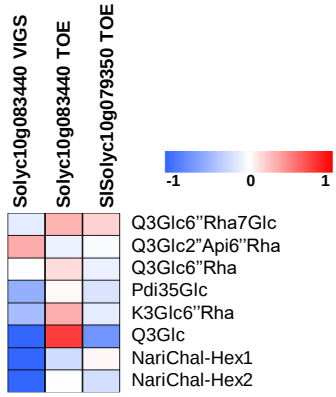
Supplemental Figure 4. Schematic of the metabolite flow amongst the flavonoid decoration steps in leaves of tomato species.

Supplemental Figure 5



SlUGT73B-a

Supplemental Figure 5. *In-silico* gene expression analysis of LYCO, PIMP, HABR and PENN seedlings for prediction of FGlcT from SlUGT genes. RNAseq data by Koening et al was used.



Supplemental Figure 6. Metabolite profiling of VIGS lines of pTRV2-Solyc10g083440 or pTRV2-Solyc10g079350. Metabolic change (fold change) is shown by a heatmap using MeV software (http://mev.tm4.org/).