the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2020) **103,** 2007–2024 doi: 10.1111/tpj.14879

Quantitative trait loci analysis of seed-specialized metabolites reveals seed-specific flavonols and differential regulation of glycoalkaloid content in tomato

Saleh Alseekh[1,2]*, Itai Ofner[3], Zhongyuan Liu[4], Sonia Osorio[5], Jose Vallarino[1], Robert L. Last[6,7], Dani Zamir[3], Takayuki Tohge[1] and Alisdair R. Fernie[1,2]*

[1]*Max-Planck-Institute of Molecular Plant Physiology, Am Mühlenberg 1, Potsdam-Golm14476, Germany,*
[2]*Center of Plant Systems Biology and Biotechnology, Plovdiv 4000, Bulgaria,*
[3]*Faculty of Agriculture, The Robert H. Smith Institute of Plant Sciences and Genetics in Agriculture at the Hebrew University of Jerusalem, Rehovot 76100, Israel,*
[4]*Horticultural Sciences, Plant Innovation Center, University of Florida, Gainesville, FL 32611, USA,*
[5]*Department of Molecular Biology and Biochemistry, Instituto de Hortofruiticultura Subtropical y Mediterranea "La Major" – University of Malaga – Consejo Superior de Investigaciones Cientificas (IHSM-UMA-CSIC), Campus de Teatinos, Malaga 29071, Spain,*
[6]*Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, USA, and*
[7]*Department of Plant Biology, Michigan State University, East Lansing, MI 48824, USA*

Received 1 April 2019; revised 5 June 2020; accepted 12 June 2020; published online 15 June 2020.
*For correspondence (e-mails alseekh@mpimp-golm.mpg.de (SA); fernie@mpimp-golm.mpg.de (AF)).

SUMMARY

Given the potential health benefits (and adverse effects), of polyphenolic and steroidal glycoalkaloids in the diet there is a growing interest in fully elucidating the genetic control of their levels in foodstuffs. Here we carried out profiling of the specialized metabolites in the seeds of the *Solanum pennellii* introgression lines identifying 338 putative metabolite quantitative trait loci (mQTL) for flavonoids, steroidal glycoalkaloids and further specialized metabolites. Two putative mQTL for flavonols and one for steroidal glycoalkaloids were cross-validated by evaluation of the metabolite content of recombinants harboring smaller introgression in the corresponding QTL interval or by analysis of lines from an independently derived backcross inbred line population. The steroidal glycoalkaloid mQTL was localized to a chromosomal region spanning 14 genes, including a previously defined steroidal glycoalkaloid gene cluster. The flavonoid mQTL was further validated via the use of transient and stable overexpression of the *Solyc12g098600* and *Solyc12g096870* genes, which encode seed-specific uridine 5′-diphosphate-glycosyltransferases. The results are discussed in the context of our understanding of the accumulation of polyphenols and steroidal glycoalkaloids, and how this knowledge may be incorporated into breeding strategies aimed at improving nutritional aspects of plants as well as in fortifying them against abiotic stress.

Keywords: flavonol, tomato, steroidal glycoalkaloids, introgression lines, seeds, specialized metabolites.

INTRODUCTION

Tomato represents an important source of fiber and nutrients in our diet and is currently the central model for fruit biology. Tomato fruits contain hundreds of structurally diverse metabolites and their consumption has been related to the reduced incidence of several chronic diseases, including cardiovascular disease (Willcox *et al.*, 2003) and certain cancers (Giovannucci, 1999). These health benefits are frequently linked to their high content in antioxidant secondary compounds, in particular flavonoids (Zhang *et al.*, 2015; Martin and Li, 2017). Over the last decade, several studies have focused on comparing the levels of these compounds among different tomato varieties (Tohge *et al.*, 2005; Moco *et al.*, 2006; Iijima *et al.*, 2008; Gomez-Romero *et al.*, 2010; Schilmiller *et al.*, 2010; Rohrmann *et al.*, 2011). The majority of flavonoids accumulate in the surface layer either in the epidermis for quercetin, kaempferol, naringenin and naringenin chalcone derivatives (Krause and Galensa, 1992; Muir *et al.*, 2001), or in the epidermis and vascular attachment region as seen for quercetin and kaempferol triglycerides (Moco *et al.*, 2006).

© 2020 The Authors.
The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

In addition to fruit quality, seed quality traits such as protein, starch and oil contents have been increasingly studied because they represent an important source of these nutrients (Bentsink *et al.*, 2000; Koornneef *et al.*, 2002; Angelovici *et al.*, 2010). However, despite the fact that the tomato fruit is widely studied, relatively little research has focused on seed secondary metabolism. Ferreres *et al.* (2010) identified 14 flavonoids in tomato seed extract using high-performance liquid chromatography/photodiode array detector/electrospray ionization-mass spectrometry (MS) and evaluated some of their biological and antioxidant capacities as well as characterizing seed-specific flavonol-*O*-sophorosides (quercetin-*3*-*O*-sophoroside, kaempferol-*3*-*O*-sophoroside and isorhamnetin-*3*-*O*-sophoroside) and their derivatives. Further studies reported the phenolic composition of the seed cavity, including the columella and placenta tissues as well as the seeds themselves (Torres *et al.*, 2005; Chandra and Ramalingam, 2011; Valdez-Morales *et al.*, 2014). Moreover, Capanoglu *et al.* (2008) studied the antioxidant activities and metabolic changes, of the major flavonoid derivatives rutin, rutin apioside, naringenin and naringenin chalcone in a mixture of seed and skin tissues. Other studies, focused on the composition of tomato seed oil (Lazos *et al.*, 1998) and the unsaponifiable matter isolated from tomato seeds (Malecka, 2002).

Steroidal glycoalkaloids (SGAs) are secondary metabolites harbored by plants of the Solanaceae family. They are abundant in tissues, including leaves, roots, flowers, fruits (e.g., in tomato and eggplant) and tubers (e.g., in potato) (Friedman and Levine, 1998; Friedman, 2002; Kozukue *et al.*, 2008; Itkin *et al.*, 2013; Sawai *et al.*, 2014; Sonawane *et al.*, 2016). Different combinations of the steroidal alkaloid aglycone and sugar moieties yield the massive structural diversity of SGAs. In tomato plants, α-tomatine, which consists of tomatidine and lycotetraose is well established as the major SGA and is found predominantly in leaves and immature fruits (Friedman, 2002; Kozukue *et al.*, 2004; Cataldi *et al.*, 2005). The major metabolic shift noted in SGA composition in tomato fruit is the ripening-dependent conversion of tomatines to esculeosides, which occurs via concerted hydroxylation and glycosylation reactions (Fujiwara *et al.*, 2007; Iijima *et al.*, 2008; Katsumata *et al.*, 2011). While the most abundant of the tomatidines, α-tomatine, is toxic to a variety of fungi, insects and animal cells, including cancer cells (Blankemeyer *et al.*, 1997; Friedman *et al.*, 2007; Milner *et al.*, 2011), esculeosides such as esculeogenin A have been reported to lead to reduce atherogenesis (Fujiwara *et al.*, 2007). Hence, SGAs have both nutritional and antinutritional roles. In plants, toxic SGAs appear to dissuade animals from eating immature fruit while by contrast the non-toxic SGAs of mature fruits potentially confer health benefits to seed-dispersing mammals (Tohge *et al.*, 2014).

Despite the identification and isolation of many secondary metabolites in tomato, their exact biosynthetic routes and molecular bases of the pathways responsible for the majority of these compounds remain unknown. In tomato, a set of 76 introgression lines (ILs) was developed by crossing the cultivated *Solanum lycopersicum* with its distant wild relative *Solanum pennellii* (Eshed and Zamir, 1995). This population has been used to study the natural variation of primary metabolites in fruit (Schauer *et al.*, 2006, 2008; Do *et al.*, 2010), leaves (Nunes-Nesi *et al.*, 2019) and seeds (Toubiana *et al.*, 2012, 2015). In this population quantitative trail loci (QTL), analyses have also been carried out on volatile metabolites, antioxidants, pigments, cell wall components, lipids and acyl sugars (Liu *et al.*, 2003; Rousseaux *et al.*, 2005; Tieman *et al.*, 2006a,b; Fraser *et al.*, 2007; Schilmiller *et al.*, 2010; Quadrana *et al.*, 2014; Garbowicz *et al.*, 2018). More than 3000 QTL associated with plant yield phenotypes and fruit metabolism have been discovered (Alseekh *et al.*, 2013). Recently, a large-scale metabolic QTL (mQTL) analysis was performed on *S. pennellii* ILs to investigate the QTL associated to specialized metabolites in tomato fruit pericarp (Alseekh *et al.*, 2015). Here we extended this study and performed secondary metabolite analysis in tomato seeds of the *S. pennellii* population. In doing so, we were able to determine the genetic architecture of seed specialized metabolites and found 338 putative mQTL for hydroxycinnamates, flavonol and glycoalkaloid compounds. Furthermore, we report seed-specific specialized metabolites and focus on two mQTL for flavonol on chromosome 12, and glycoalkaloid on chromosome 7. Both mQTL were validated using the *S. pennellii* backcross inbred lines (BILs) (Ofner *et al.*, 2016) and sub-ILs. In addition, both stable transgenic and transient overexpression lines were used to validate two candidate genes underlying the seed-specific flavonol. The occurrence of this compound was additionally studied across a range of wild species of the lycopersicum complex (Schauer *et al.*, 2005), as well as several different accessions of *S. pennellii* with results suggesting a possible ecological role for these compounds. These combined results are discussed in the context of our current understanding of flavonol and glycoalkaloid contents in tomato as well as in more general terms with regard to genetic analysis of plant metabolism.

RESULTS

Identification of QTL for seed secondary metabolites in the *S. pennellii* IL population

To identify potential QTL involved in the regulation of secondary metabolites levels in tomato seeds, we used a set of 76 ILs derived from a cross between the domesticated *S. lycopersicum* (cv M82) and its wild relative *S. pennellii*

© 2020 The Authors.

(Eshed and Zamir, 1995). The IL population is composed of 76 ILs distributed across the 12 tomato chromosome, each containing a single introgression from *S. pennellii* (LA 716) in the genetic background of the processing tomato variety M82, and cover the full *S. pennellii* genome. Methanol extracts of dry seeds of the ILs were subjected to liquid chromatography (LC)-MS analysis for metabolite profiling. In total, we were able to quantify 74 secondary metabolites, including 35 annotated and 39 unknown compounds (Tables S1 and S2). Figure 1 shows a heat map of the metabolite profiling results, where red and blue rectangles depict increases and decreases with respect to the average of all ILs and the parental M82, respectively. The relative differences in the content of any given metabolite range between 0 (absent; as seen for quercetin-*3*-*O*-sophoroside-*O*-rhamnoside (ID: M20), kaempferol-*3*-*O*-sophoroside-*O*-rhamnoside (ID: M58) and isorhamnetin-*3*-*O*-sophoroside-*O*-rhamnoside (ID: 23) and a 510-fold increase (in isorhamnetin-*3*-*O*-sophoroside-*O*-rhamnoside (ID: 56) compared with the parent variety M82 (Figures 1 and 2a).

Next, data from 74 metabolites present in 76 ILs were used to map QTL putatively using the analysis of variance test, which was performed at two levels of significance hereafter referred to as permissive ($P \leq 0.05$) and stringent ($P \leq 0.01$). In total, 801 putative mQTL and 338 putative mQTL were identified at the permissive and stringent levels, respectively (Figure S1). Given the vast number of putative QTL from the permissive test, we decided to focus only on the stringent level. The majority of the QTL found were increased in the ILs with only 20% decreased with respect to M82. While it is possible that this is due to fact it is difficult to measure decreases in low-level metabolites, it is interesting to note that such a finding was also observed in the fruit (Alseekh *et al.*, 2015, 2017) and that *S. pennellii* is anticipated to contain higher levels of secondary metabolites (Schwahn *et al.*, 2014; Tohge *et al.*, 2020). In total, 184 of these QTL were for 27 of the annotated metabolites, which included flavonoids and glycoalkaloids (Tables S1 and S2). The metabolites with the most complex genetic architecture, i.e., having the most QTL were kaempferol-*3*-*O*-rutinoside (19 QTL), (caffeoyl)rutinoside-*5*-*O*-glucoside (17 QTL), dicaffeoylquinic acid (11 QTL), quercetin-hexose 1 (11 QTL), eriodictyol-hexose (10 QTL) and -hexose I (9 QTL). In addition, the glycoalkaloid derivatives, m/z = 1081.6, m/z = 1198.6 and m/z = 933.4, showed significant changes in several ILs compared with M82 (17, 13

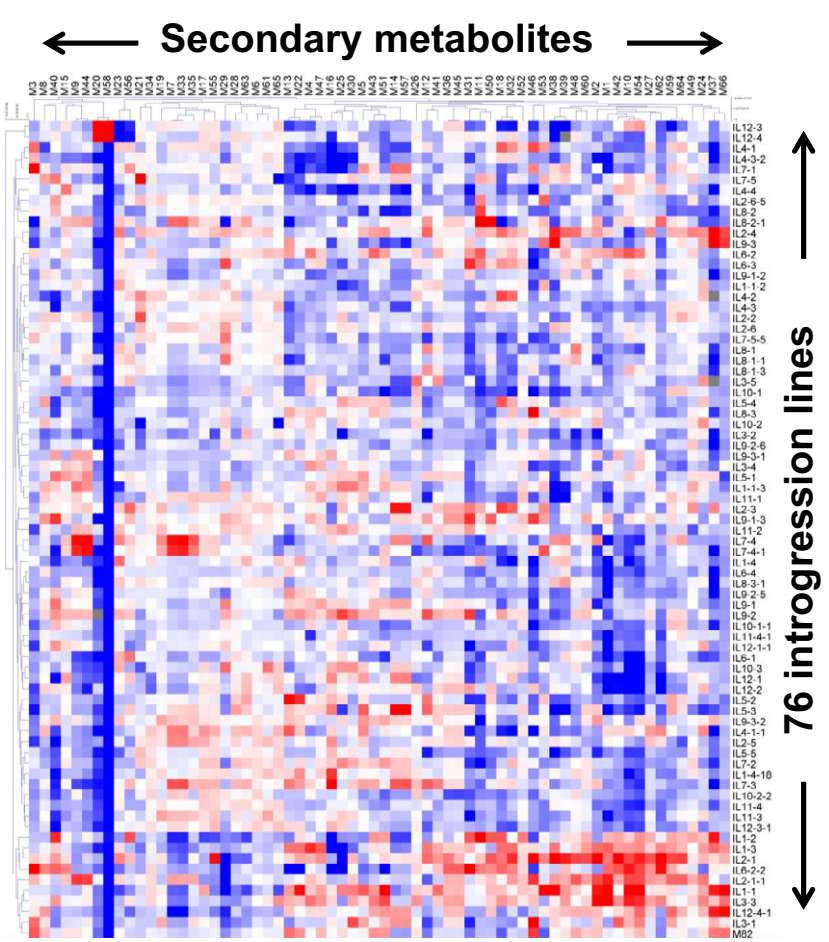


Figure 1. Hierarchical clustering heat map representation of changes in secondary metabolite levels measured on dry introgression lines seeds. Regions of red or blue indicate that the metabolite content is increased or decreased, respectively, after the introgression of *S. pennellii* segments.
Values presented are means of four replicates and are shown in false-color code. The metabolite annotation and complete data set are presented in Table S1.

© 2020 The Authors.
The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd,

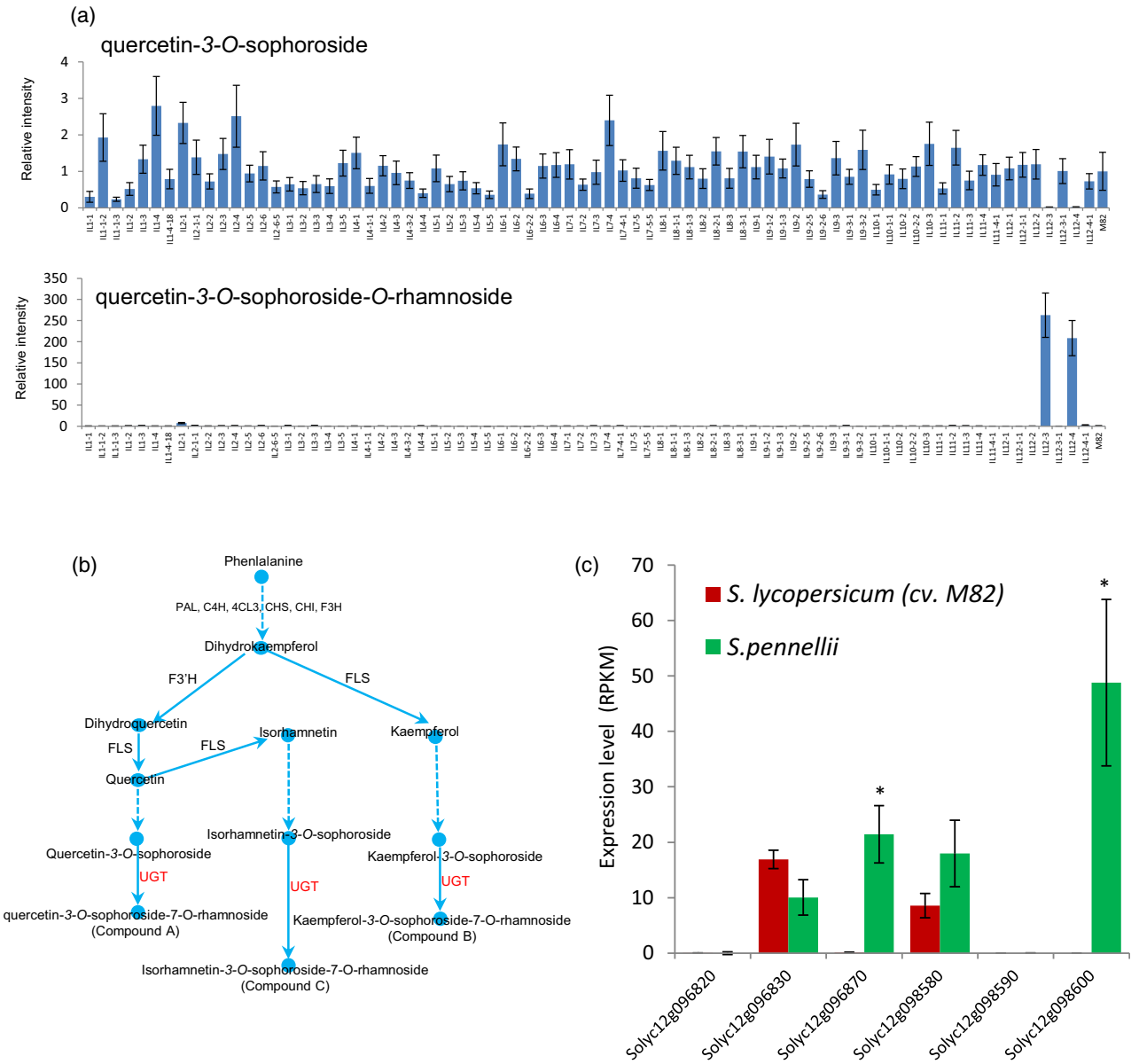


Figure 2. Metabolite and expression levels of flavonol associated metabolism.
(a) Levels of major detected compound (quercetin-*3-O*-sophoroside-*O*-rhamnoside) and its upstream precursor (quercetin-*3-O*-sophoroside) in the dry seeds of *S. pennellii* introgression lines. Values presented are the mean of four to five biological replicates ± SE.
(b) Schematic representation of the flavonoid biosynthetic pathway. 4CL, 4-coumarate, coenzyme A ligase; C3H, 4-coumarate 3-hydroxylase; C4H, cinnamate 4-hydroxylase; CHI, chalcone isomerase; CHS, chalcone synthase; DFR, dihydroflavonol reductase; F3′5′H, flavonoid-3′5′-hydroxylase; F3′H, flavonoid-3′-hydroxylase; F3H, flavanone-3-hydroxylase; FLS, flavonol synthase; PAL, phenylalanine ammonia lyase; UGT, UDP-glycosyltransferase. Also showing are the major detected compounds in this study.
(c) Expression levels of the six UGT genes (*Solyc12g096820, Solyc12g098590, Solyc12g096830, Solyc12g098580, Solyc12g096870, Solyc12g098600*) within the metabolite quantitative trait loci region. Data from RNA-seq of *S. lycopersicum* (cv M82) and the *S. pennellii* wild species (Bolger *et al.*, 2014).
Values are provide in RPKM represent the mean of three to five biological replicates ± SE. Asterisks indicate significant differences from the control (M82) as assessed by Student's *t*-test ($P < 0.01$). *$P \leq 0.01$.

and 16 QTL respectively). By contrast, other metabolites displayed significant changes in a single or only a few specific lines, such as caffeoyl-hexoside I, which was 4.3-fold increased or decreased in IL1-3 compared with M82, lycoperoside A, B or C, which was 11.6-fold increased or decreased in IL2-1 compared with M82, rutin, which was increased or decreased by 4.4 and 0.26 in ILs 1-2 and 12-4 compared with M82, respectively; chlorogenic acid was increased or decreased in ILs 1-1-3 and 7-3, respectively; and esculeoside A, or lycoperoside G or F were increased

© 2020 The Authors.

or decreased in ILs 1-1, 2-1 and 9-3 by 6.7, 6.1, 0.1 compared with M82, respectively. A more detailed analysis of the metabolite changes is presented in Table S1. Furthermore, the overlapping ILs 12-3 and 12-4 showed specific changes in four flavonoid compounds; quercetin-3-*O*-sophoroside-*O*-rhamnoside (771.2 *m*/*z,* Figure 2a), kaempferol-3-*O*-sophoroside-*O*-rhamnoside (755.3 *m*/*z*), isorhamnetin-3-*O*-sophoroside-*O*-rhamnoside (785.2 *m*/*z*) and quercetin diglycosides (625.2 *m*/*z*). The distribution of the putative seed mQTL across the tomato genome is represented in Figure S1. The results showed that the seed mQTL are unevenly distributed and there are several hotspots in which more mQTL are detected, particularly loci on chromosome 1 (IL1-1, IL1-2), 2 (IL2-1, IL2-3, IL2-5), 6 (IL6-2-2), 7 (IL7-3) and 9 (IL9-2).

Fine mapping of candidate genes of the mQTL for seed-specific flavonoids

We next adopted a combination of genetic mapping, genomics and biochemical approaches for the identification of genes underlying the changes in metabolite accumulation found in the tomato seeds. QTL analysis of tomato ILs revealed that seeds produce specific flavonoid glycosides with quercetin, kaempferol and isorhamnetin in aglycone moieties (Figure 3b–f). Comparisons of LC-MS total-ion chromatograms for M82 and those from ILs 12-3 and 12-4 with introgressed overlapping from *S. pennellii* revealed clear changes in three and two major peaks from the chromosomal substitution lines and M82, respectively (Figure 3b–f). The mass spectra in negative mode of the observed compounds are listed according to their elution time from early to late elution: A, *m*/*z* = 771.2 (Figure 3b); B, *m*/*z* = 755.3 (Figure 3c); and C, *m*/*z* = 785.2 (Figure 3d). These peaks were only observed in extracts from ILs 12-3 and 12-4, which have overlapping introgressed segments from *S. pennellii* LA0716. On the other hand the LC-MS chromatograms showed two highly abundant compounds in the cultivated variety M82. The parent ion mass of these compounds are, respectively, listed according to their elution time from early to late elution: D, *m*/*z* = 625.1 (Figure 3e); and E, *m*/*z* = 609.1 (Figure 3f). Together the mass spectra indicated the presence of flavonoid-*O*-*tri*-glycosides for compounds A, B and C, and the source fragmentation of compounds A, B and C, showed that the aglycone-related ions are respectively *m*/*z* = 300, quercetin; *m*/*z* = 285, kaempferol; and *m*/*z* = 315, isorhamnetin. In addition, the fragmentation profiles of these peaks displayed unique ions corresponding to neutral loss of a deoxyhexose (rhamnose) residue (146 u). When taken together, we annotated these peaks putatively as quercetin-*3*-*O*-sophoroside-*O*-rhamnoside (compound A), isorhamnetin-3-*O*-sophoroside-*O*-rhamnoside (B) and kaempferol-*3*-*O*-sophoroside-*O*-rhamnoside (C). By contrast, peaks D (*m*/*z* = 625.1) and E (*m*/*z* = 609.1) accumulated to considerable levels in M82 but were undetectable or negligibly present in the overlapping ILs 12-3 and 12-4 (Figure 3e,f). The fragmentation and accurate mass measurements indicated that these peaks are isorhamnetin-*3*-diglucoside and quercetin-*3*-rutinoside (rutin), respectively. Furthermore, combining information from accurate measurements and knowledge of flavonoid biosynthesis revealed a loss of glycosylation in *S. lycopersicum* (cv M82), which is characteristic of flavonol-3-*O*-glycoside-*O*-rhamnoside. The results of QTL mapping demonstrate that the putative mQTL for the loss of glycosylation mapped to the overlapping regions of ILs 12-3 and 12-4 on chromosome 12 (Figures 2a and S1). We hypothesized that the overlapping region likely harbors one or more flavonoid rhamnosyltransferase genes that are functional in *S. pennellii* but are either absent or non-functional in *S. lycopersicum* (cv M82).

We next used the genetic markers that define the ends of the overlapping region of IL12-3 and IL12-4 on chromosome 12 to identify the genomic sequence in the assembly of tomato genome (Tomato Genome, 2012), available at the Sol genomics network (assembly ITAG3.0; http://solgenomics.net). This region was predicted to contain 344 genes, including nine annotated as uridine 5′-diphosphate (UDP)-glycosyltransferase (*Solyc12g088700*, *Solyc12g088710*, *Solyc12g096080*, *Solyc12g096820*, *Solyc12g096830*, *Solyc12g096870*, *Solyc12g098580*, *Solyc12g098590* and *Solyc12g098600*). Checking the *S. pennellii* genome in this region revealed one additional gene with unknown function; however, this gene appears not to be expressed in the seed and is not a rhamnosyltransferase. To narrow down the number of candidate genes and validate our finding in a different mapping population, we took advantage of available backcross inbred lines (BILs) (Ofner *et al.*, 2016) and a set of sublines (Garbowicz *et al.*, 2018) to analyze the secondary metabolites in seeds from these lines. Figure 3 shows the changes in the levels of five flavonol glycosides in the seeds of 16 sublines derived from the cross of IL12-3 to M82 (Figure 3). By comparing the flavonoid glycoside content of each sub-IL to M82, we mapped the changes in the levels of these compounds to a region defined by the molecular markers C201757 and S422-1. Five sublines showed higher accumulation of three major flavonoid glycosides: quercetin-*3*-*O*-sophoroside-*O*-rhamnoside, isorhamnetin-*3*-*O*-sophoroside-*O*-rhamnoside and kaempferol-*3*-*O*-sophoroside-*O*-rhamnoside (Figure 3b–d). By contrast, the levels of kaempferol-*3*-*O*-sophoroside and isorhamnetin-*3*-*O*-sophoroside were substantially reduced in another 11 mapping sublines (3128-72-53, 3100-10, 3102-2, 3049-11-29, 3049-25-109, 3049-164-16, 3148-76-31, 3113-11-1, 3166-2-77, 3128-133-44, 3128-106-B223) sharing the same genomic region from the cultivated variety and had substantially higher levels of flavonoid diglycosides and undetectable or

© 2020 The Authors.
The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd,

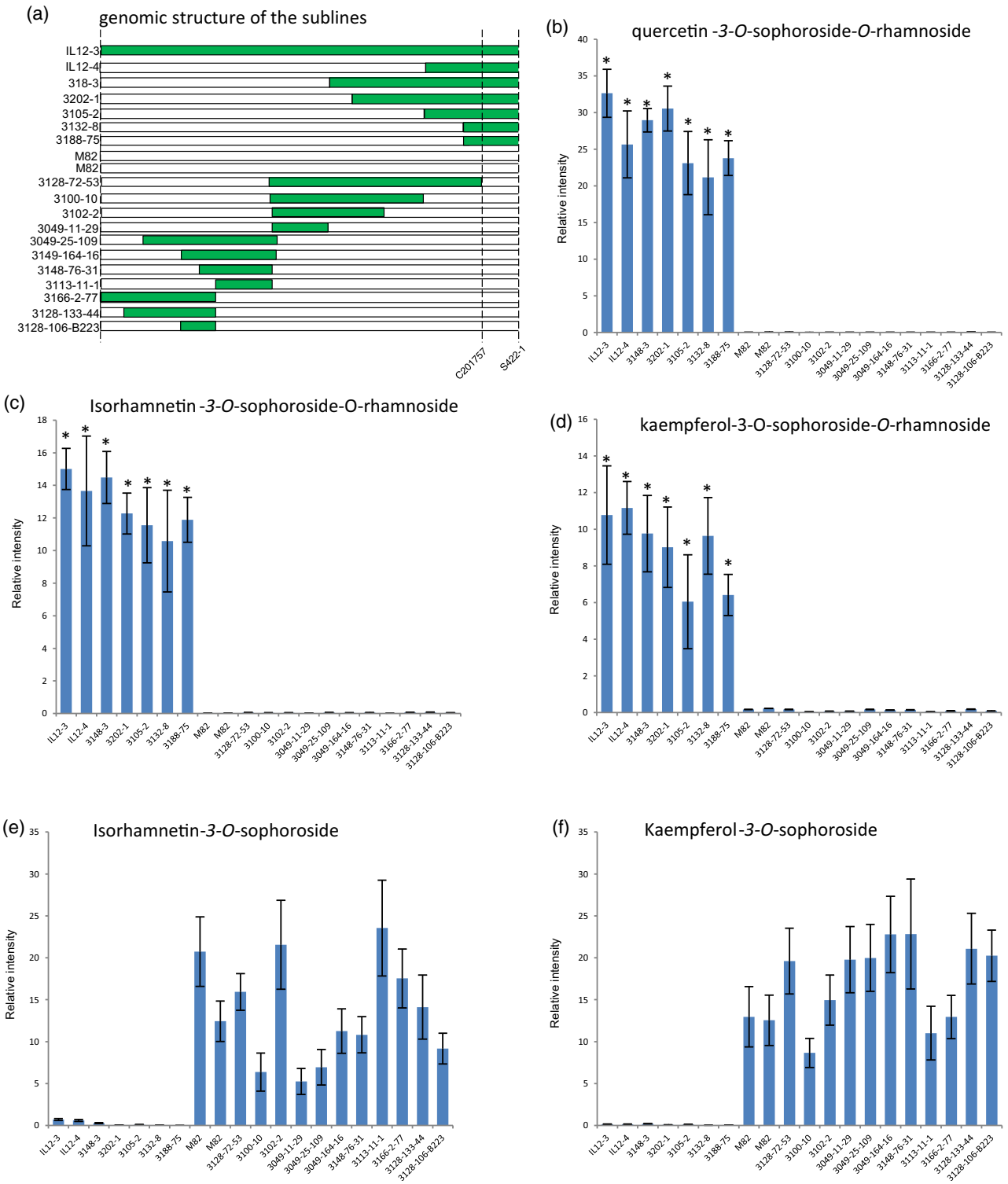


Figure 3. Sub-introgression line (IL)-based fine mapping of flavonol derivatives.
(a) Showing genomic structure of the sublines generated from IL 12-3, black bars indicate *S. pennellii* like, white bars indicate M82 like. Through the seed analysis of 16 lines, five sublines (318-3, 3202-1, 3105-2, 3132-8, 3188-75) displayed the same phenotypes with ILs 12-3 and 12-4.
(b) Quercetin-*3-O*-sophoroside-*O*-rhamnoside, (c) isorhamnetin-*3-O*-sophoroside-*O*-rhamnoside, (d) kaempferol-*3-O*-sophoroside-*O*-rhamnoside, (e) isorhamnetin-*3-O*-sophoroside, (f) kaempferol-*3-O*-sophoroside. Values are presented as the mean ± SE of three to five biological replicates.

© 2020 The Authors.

neglected amounts of flavonoid triglycosides (Figure 3e–f). The annotation of the tomato genome between the defined markers indicates the presence of approximately 160 genes. To validate our finding further and to narrow down the number of candidate genes, we performed LC-MS analysis in dry seeds of 15 BILs (Ofner *et al.*, 2016) covering the overlapping region of ILs 12-3 and 12-4 (Figure 4). These BILs were created by different breeding crosses compared with *S. pennellii* ILs and the subset of sublines, but sharing the same parental lines. The metabolic profiling of 15 BILs indicated that eight of 15 lines showed very similar metabolite profiles to IL12-4. These lines all share the same *S. pennellii* sequence and are predicted to contain about 60 genes, including six UDP-glycosyltransferases (UGT). Surveying the *S. pennellii* genome revealed a highly similar gene density in this region (Bolger *et al.*, 2014). We next checked the gene expression of all candidates in this region in RNA-seq data available from M82 and *S. pennellii* (Bolger *et al.*, 2014) (Figure 2c). The results showed that there were no expressions for two candidate genes, *Solyc12g096820* and *Solyc12g098590*. Two further genes, *Solyc12g096830* and *Solyc12g098580*, had no significant changes in expression between the M82 and *S. pennellii*. Interestingly, however, two UGTs, *Solyc12g096870* and *Solyc12g098600*, were highly expressed and showed six- and 150-fold higher expression in *S. pennellii* compared with cv M82. To validate these results, gene-specific polymerase chain reaction (PCR) primers of all six candidates were used to check gene expression using cDNA prepared from fresh seeds extracted from IL12-3, IL12-4 and M82. Two of the candidates, *Solyc12g096870* and *Solyc12g098600*, showed the presence of mRNA only in IL12-3 and IL12-4 seeds, while no differences in the expression were found for the other candidates (Figure S2a). For those genes displaying invariant expression, we compared the sequence of their coding regions as further support for the observed phenotypes. We performed genomic sequence analysis of the promoter regions (defined as 1000 bp upstream of start codon) of both *Solyc12g096870* and *Solyc12g098600* and compared the promoter sequences by aligning the sequences from *S. lycopersicum* cv M82 and *S. pennellii* genome sequences (Bolger *et al.*, 2014). Several deletions or insertions in the M82 promoter sequences of both genes compared with the *S. pennellii* promoter sequences are deleted (Figure S3). By contrast, the coding regions were very similar between M82 and *S. pennellii* with five and four amino acids changes between *S. lycopersicum* cv M82 and *S. pennellii* in *Solyc12g096870* and *Solyc12g098600*, respectively (Figure S4a,b). On the other hand, the other genes showed nine, four, five and 42 amino acid changes *S. lycopersicum* cv M82 and *S. pennellii* for *Solyc12g096820*, *Solyc12g098590*, *Solyc12g096830* and *Solyc12g098580*, respectively. Taken together, these results suggest that the wild-type alleles of one or both candidate genes have elevated expression, resulting in accumulation of flavonoid triglycosides in *S. pennellii* compared with the cultivated variety.

Functional validation of candidate genes *Solyc12g096870* and *Solyc12g098600*

Next, we subjected the UGTs (*Solyc12g096870* and *Solyc12g098600*) to phylogenetic analysis to determine their grouping patterns and genetic relationships relative to characterized genes encoding flavonoid UGTs in other plant species (Figure S5). The phylogenetic analysis showed three clear groups according to their enzymatic functions (3GT, 5GT and 7GT). Both candidate genes (*Solyc12g096870* and *Solyc12g098600*) were clustered within the group of 5GTs albeit also closely to the 7GTs. These results suggest that *Solyc12g096870* and *Solyc12g098600* act as either 5GT or 7GT in tomato seeds.

To test whether either or both *Solyc12g096870* and *Solyc12g098600* UGTs were involved in the metabolic changes observed, we next carried out targeted reverse genetics experiments. We generated 35S stable overexpression transgenic plants expressing the *S. pennellii* alleles of *Solyc12g096870* and *Solyc12g098600* in the cultivated background. In total, 21 and 20 independent stable transgenic overexpressing lines were subjected to LC-MS analysis for UGT-*Solyc12g098600* and UGT-*Solyc12g096870*, respectively (Figure S6). Data indicated that several independent overexpression lines of the UGT-*Solyc12g098600* have higher levels of quercetin-*3-O*-sophoroside-*O*-rhamnoside, kaempferol-*3-O*-sophoroside-*O*-rhamnoside and isorhamnetin-*3-O*-sophoroside-O-rhamnoside compared with the control. By contrast, there were no obvious significant differences in the levels of these compounds between the transgenic plants overexpressing *Solyc12g096870* and the wild-type plants. For further validation, we transiently overexpressed both UGT-*Solyc12g098600* and UGT-*Solyc12g096870* genes by infiltrating tobacco leaves with agrobacteria containing the respective overexpression constructs (Figure S6). The results revealed a more than threefold increase in flavonol-*3-O*-triglycosides in leaf inoculated with UGT-*Solyc12g098600* compared with an empty vector, and on the other hand showed only a 1.6-fold increase in leaf inoculated with UGT-*Solyc12g096870* compared with empty vector.

Flavonol levels in tomato seed, peel and fruit pericarp from different developmental stages

To gain more insight into accumulation of the observed specific flavonol in tomato seeds, we grew M82 alongside ILs 12-3 and 12-4, and harvested fresh seeds from fruit at four different developmental stages: green (35 days after pollination [DAP]), breaker_1 (45 DAP), breaker_2 (50 DAP) and red ripe (55 DAP). LC-MS analysis revealed that quercetin-*3-O*-sophoroside-*O*-rhamnoside

© 2020 The Authors.
The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd,

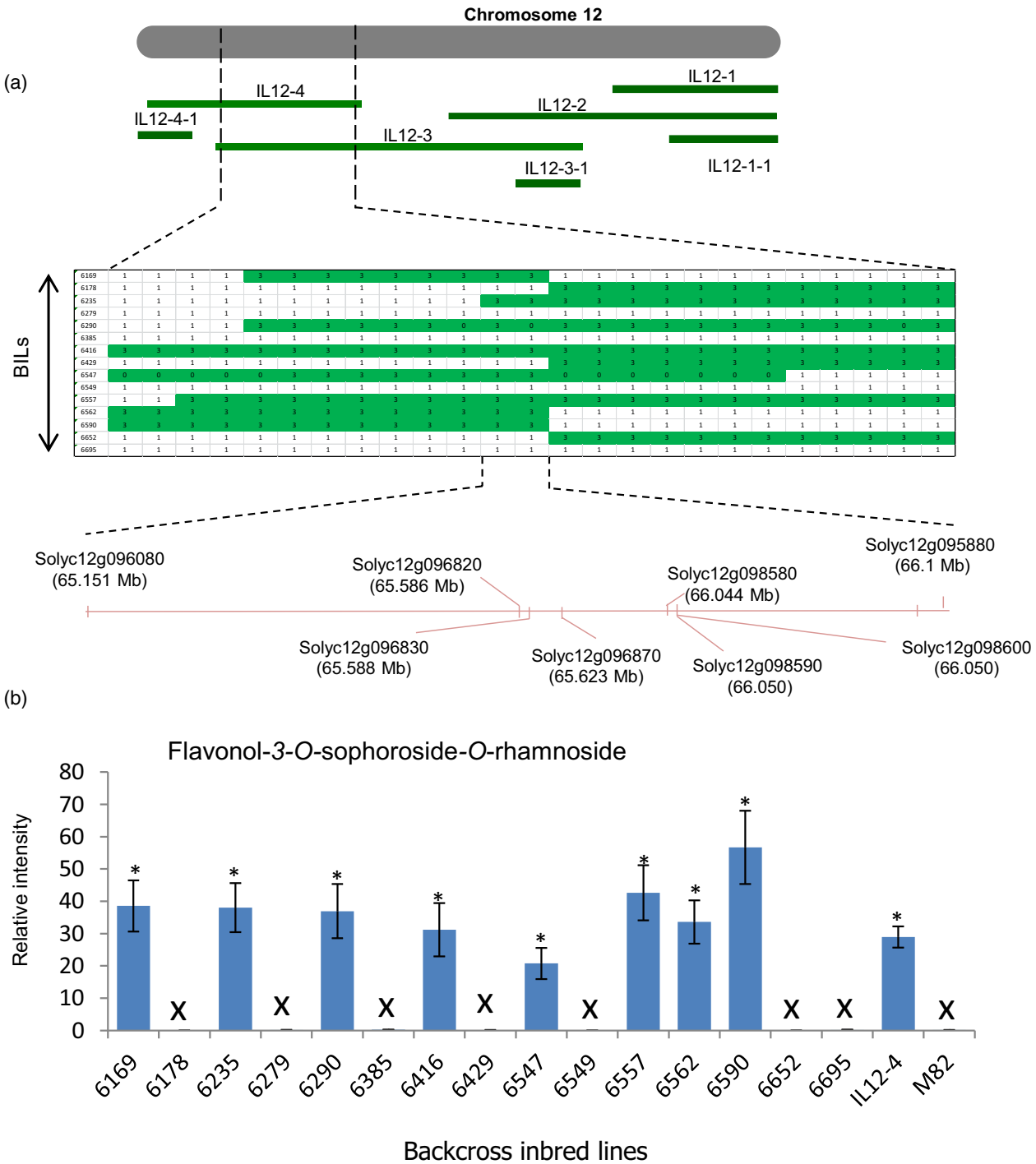


Figure 4. Backcross inbred line (BIL)-based fine mapping of flavonoid metabolite quantitative trait loci.
(a) Showing genomic structure of the BILs used for fine mapping of the metabolite quantitative trait loci, green bars with "3" indicates the markers showing *S. pennellii* like, white bars with "1" indicates markers with M82 like. Through the analysis of 15 BILs, 8 BILs; 6169, 6235, 6290, 6416, 6547, 6557, 6562 and 6590 showed same phenotypes with introgression line (IL)12-4. Figure also showing a magnified view of candidate genes.
(b) Levels of quercetin-3-O-sophoroside-O-rhamnoside in BILs, IL12-4 and M82 seeds. Values are presented as the mean ± SE of three to five biological replicates. Asterisks indicate significant differences from the control (M82) as assessed by Student's *t*-test ($P < 0.001$). *$P \leq 0.001$. x, below the levels of detection.

dramatically increased in the seeds collected from breaker_1, breaker_2 and red ripe fruit compared with seeds collected from green stage (Figure 5), while negligible or undetectable amount were found in M82 seeds at any of the four developmental stages. The expression levels of *Solyc12g098600* and *Solyc12g096870*

© 2020 The Authors.

(Figure S2b) showed higher expression in both IL12-3 and IL12-4 in the seeds at breaker-1, beaker-2 and red. This suggested that the levels of this flavonols are associated with seed maturation and development and might involve in seed protection or dispersal. To check whether these compounds accumulated in other fruit tissues, we measured the levels quercetin-*3*-*O*-sophoroside-*O*-rhamnoside and gene expression in peel, fruit pericarp and the seeds in both ILs 12-3 and 12-4, and M82 (Figures 5b and S2a). The results indicated that negligible or none of this flavonol was detected in the peel or the fruit pericarp (Figure 5b) with higher gene expression of *Solyc12g098600* and *Solyc12g096870* in seed tissue (Figure S2b,c). This result supports our contention that the flavonol derivatives observed in this study are seed-specific compounds.

Levels of quercetin-*3*-*O*-sophoroside-*O*-rhamnoside in seeds of different tomato wild species

To assess natural variation in the levels of the observed flavonol derivatives, we quantified the levels of quercetin-*3*-*O*-sophoroside-*O*-rhamnoside in the seeds of seven wild species (*S. pennellii* [LA0718], *S. chmielewskii*, *S. chilense*, *S. lycopersicoides*, *S. neorickii*, *S. peruvianum*, *S. pimpinellifolium*) and three cultivated tomato varieties (M82, Microtom and Moneymaker) (Figure 6a). The results showed that quercetin-3-*O*-sophoroside-*O*-rhamnoside specifically accumulated in the wild species *S. pennellii* (LA0716) compared with other wild species and the cultivated varieties. This result suggested that quercetin-3-*O*-sophoroside-*O*-rhamnoside is a seed- and *S. pennellii*-specific compound. We tested whether this compound is specific to the wild *S. pennellii* (LA0716) accession or present in other *S. pennellii* accessions, quercetin-*3*-*O*-sophoroside-*O*-rhamnoside was quantified in the seeds of 25 *S. pennellii* accessions previously collected across Peru and provided by The C.M. Rick Tomato Genetics Resource Center (TGRC) (Figure 6b). Our results indicate occurrence of this metabolite is found across *S. pennellii*. However, the data also revealed that there is a high variance across the accessions (Figure 6b). For example, most of the southern *S. pennellii* accessions, including LA0716, contained six times higher levels of quercetin-3-*O*-sophoroside-*O*-rhamnoside compared with the far northern *S. pennellii* accessions (Table S4). In contrast, medium to low levels were found in the *S. pennellii* accessions located in the middle part of the country.

Fine mapping of mQTL involved in SGA accumulation

Our mQTL analysis revealed that the overlapping region of IL7-4 and IL7-4-1 on chromosome 7 is correlated with significantly increased levels of three distinct glycoalkaloids peaks at m/z = 1197.7, m/z = 1065.9 and m/z = 1067.8, compared with M82 and the other ILs (Figures 7b and S8a–e). The accurate mass and fragmentation profiles of these peaks showed a unique ion corresponding to the loss of a hexose residue (162 mass unite). In addition, these peaks showed very similar mass spectra, leading us to hypothesize that the putative candidate genes encode one or more UGT, functional in *S. pennellii* and either absent from or non-functional in *S. lycopersicum* (cv M82).

The overlapping genomic region of IL7-4 and IL7-4-1 ILs on chromosome 7 (Figure 7a) was used to identify the genomic sequence in the tomato genome assembly (Sol genomics network, assembly ITAG3.1; http://solgenomics.net). This region is predicted to contain 1072 genes, including 19 annotated as UGTs. Following the same approach as described above, we used a subset of BILs covering the overlapping region of ILs 7-4 and 7-4-1 to narrow down our mQTL region and the number of candidate genes. For this purpose, we conducted LC-MS metabolic profiling targeting the main peaks in seeds from 10 BILs (Figure 7b,c). Four lines (6428, 6489, 6612 and 6703) exhibited significantly higher levels of glycoalkaloid derivatives (m/z = 1197.7, m/z = 1065.9 and m/z = 1067.8) similar to the phenotype observed in IL7-4-1. In contrast, low levels of glycoalkaloid derivatives were found in six BILs (6192, 6222, 6281, 6415, 6548 and 6552) containing the cultivated variety genomic background region, respectively. This experiment allowed us to validate our observed mQTL in the ILs, and to narrow down the mQTL region to a small genomic region spanning about 25 kb and containing 14 annotated gene models (Figure 7a). Among the annotated genes are four UGTs in the duplicated genomic regions (*Solyc07g043410*, *Solyc07g043480*, *Solyc07g043490* and *Solyc07g043500*), in addition to two genes (*Solyc07g043420*: 2-oxoglutarate-dependent dioxygenase and *Solyc07g043460*: cytochrome P450) previously reported to be associated with the SGA biosynthesis pathway in tomato (Itkin *et al*, 2013). Next, gene expression using the cDNA from seeds collected from M82, IL7-4 and IL7-4-1 and revealed higher accumulation of mRNA of *Solyc07g043410* in IL7-4 and IL7-4-1 compared with M82, while no significant changes were observed found for the other candidates within this region (Figure S9). However, four genes (*Solyc07g043410*, *Solyc07g043490*, *Solyc07g043420* and *Solyc07g043460*) were found to be significantly more highly expressed in the leaves of overlapping ILs 7-4 and 7-4-1 (Bolger *et al*., 2014). To test whether this QTL also affects other tomato tissues, we grew the ILs 7-4 and 7-4-1 and the cultivated variety and harvested leaves, fruit pericarp, peel and seeds and analyzed their specialized metabolite content via LC-MS. These experiments revealed a higher accumulation of glycoalkaloid derivatives in the seed and leaves of ILs 7-4 and 7-4-1 compared with M82 (Figures 7b and S8). By contrast, no significant changes were found in the levels of these metabolites in fruit pericarp and peel.

© 2020 The Authors.
The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd,

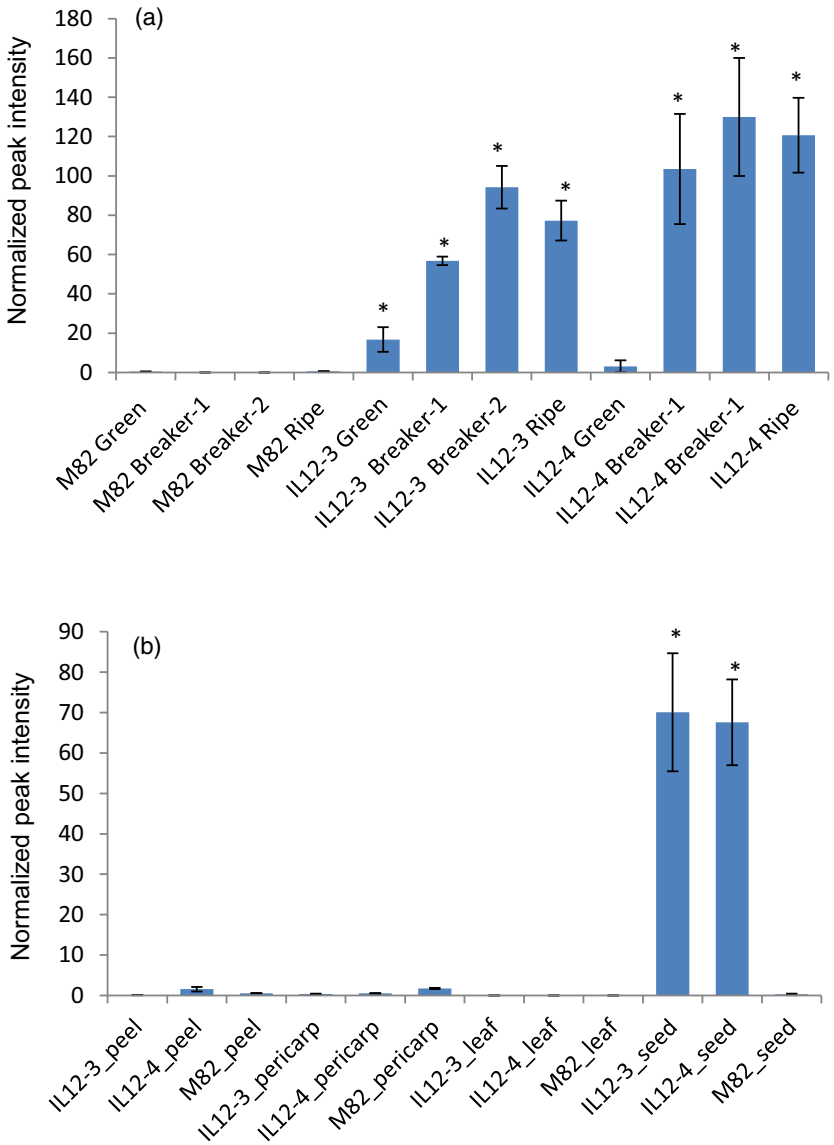


Figure 5. Tissue levels of quercetin-*3*-*O*-sophoroside-*O*-rhamnoside.
(a) Levels of quercetin-*3*-*O*-sophoroside-*O*-rhamnoside in fresh seeds collected from lines introgression line (IL)12-3, IL12-4 and cultivated variety M82 at four developmental stages; green, breaker-1, breaker-2 and red ripe fruit (35, 45, 50, 55 days after pollination) respectively.
(b) Levels of quercetin-*3*-*O*-sophoroside-*O*-rhamnoside in peel, pericarp, leaf and seed of IL12-3, IL12-4 and the cultivated variety M82. Values are presented as the mean ± SE of three to five biological replicates. Asterisks indicate significant differences from the control (M82 at each stage) as assessed by Student's *t*-test ($P < 0.01$). *$P \leq 0.01$.

DISCUSSION

Defining the genetic and biochemical regulation of seed secondary metabolite levels is important due to their roles in protection from environmental stress and plant herbivore/plant pathogen interactions (Debeaujon *et al.*, 2003; Routaboul *et al.*, 2006; Tohge *et al.*, 2016). In the current study, the genetic basis of natural variability in seed-specialized metabolism was investigated. To accomplish this aim, we initially employed metabolite profiling of seeds from tomato ILs containing segmental substitutions of the wild species in the genetic background of the cultivar M82. Most previous studies in tomato, focused on identifying mQTL in fruits and leaf tissues (Fridman *et al.*, 2004; Schauer *et al.*, 2006, 2008; Stevens *et al.*, 2007; Do *et al.*, 2010; Maloney *et al.*, 2010; Schilmiller *et al.*, 2010, 2012, Ruan *et al.*, 2012; Quadrana *et al.*, 2014; Nunes-Nesi *et al.*, 2019). Few studies have investigated the regulation of seed metabolism in tomato and those published studies focused on seed primary metabolism (Toubiana *et al.*, 2012, 2015). In the current study, we identified 338 putative mQTL for 55 specialized metabolite traits, including flavonoids, glycoalkaloids and unknown compounds in

© 2020 The Authors.

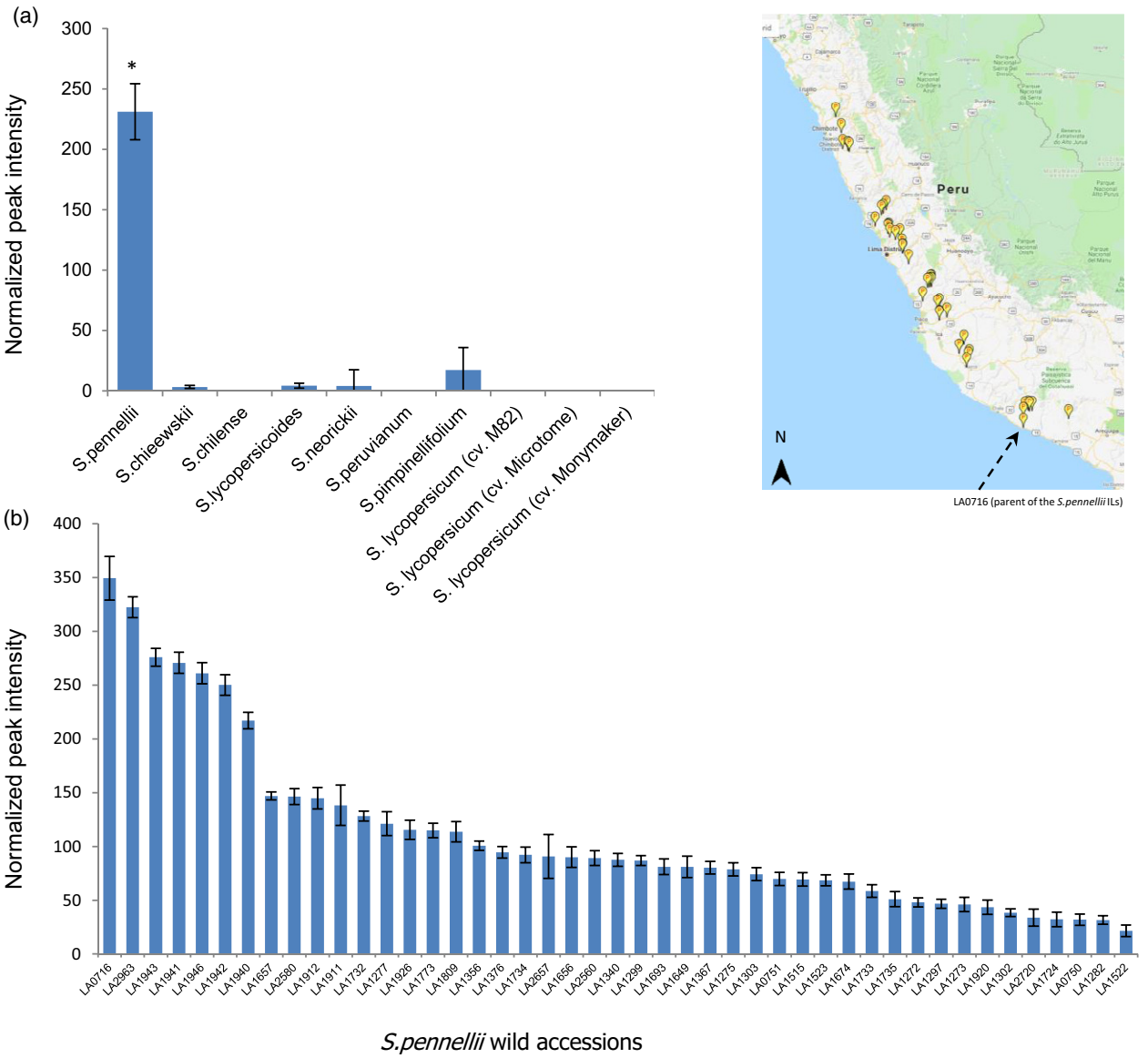


Figure 6. Flavonols levels in wide diversity of tomato species.
(a) Levels of quercetin-*3*-*O*-sophoroside-*O*-rhamnoside in dry seeds of three cultivated tomato *S. lycopersicum* (cv M82), *S. lycopersicum* (cv Microtome) *S. lycopersicum* (cv Moneymaker) and seven wild species; *S. pennellii* (LA0718), *S. chmielewskii, S. chilense, S. lycopersicoides, S. neorickii, S. peruvianum, S. pimpinellifolium*.
(b) Levels of quercetin-*3*-*O*-sophoroside-*O*-rhamnoside in dry seeds of *S. pennellii* wild accessions and their geographical origin according to The C.M. Rick Tomato Genetics Resource Center (TGRC) (Table S3), values are sorted from high to low abundance.
Values are presented as the mean ± SE of three to five biological replicates. Asterisks indicate significant differences from the control (M82) as assessed by Student's *t*-test ($P < 0.01$). *$P \leq 0.01$.

tomato seeds (Figure S1 and Table S1). On average, we observed 4.56 mQTL per specialized metabolite. When comparing these results with those from fruit secondary metabolites from the same population (Alseekh *et al.*, 2015), a similar number of 4.68 mQTL per metabolite was obtained. However, this number was considerably higher than that observed for seed primary compounds 0.48 mQTL per compound in this population (Toubiana *et al.*, 2012). Our results showed that the absolute variation in metabolites abundance across the population is much greater for secondary metabolites, which exhibited increases of up to 500-fold and decreased down to 0 (absent) of the level found in the recurrent parent M82. Furthermore, some of the metabolites showed very low coefficients of variation (<0.3), indicating that the levels of these metabolites were mainly under the influence of heritable factors. Having identified such a large number of metabolites, we chose to focus on those for quercetin-*3*-

© 2020 The Authors.

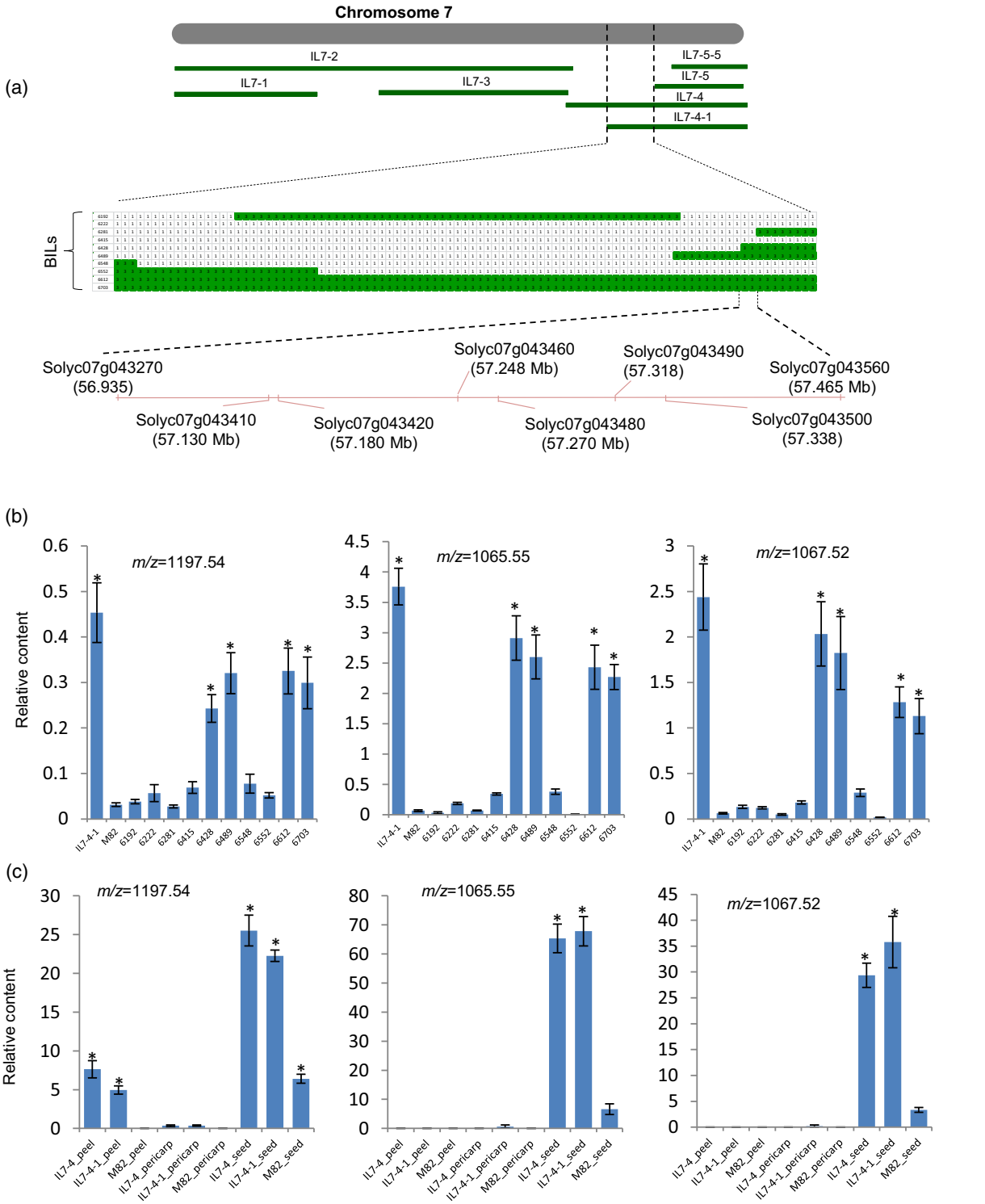


Figure 7. Fine mapping of glycoalkaloid metabolite quantitative trait loci.
(a) Showing genomic structure of the backcross inbred lines (BILs) used for fine mapping of metabolite quantitative trait loci, green bars with number indicates *S. pennellii* like, white bars indicates M82 like also showing magnified view showing the candidate genes. Through the analysis of 11 lines, four recombinant lines; 6428, 6489, 6612, 6703 showed same phenotypes with introgression line (IL)7-4-1.
(b) Levels of glycoalkaloid derivatives in control (M82), IL7-4-1 and 10 BIL seeds; m/z = 1197.7, m/z = 1065.9, m/z = 1067.8.
(c) Levels of glycoalkaloid derivatives in peel, pericarp and fresh seeds collected from tomato fruits of IL7-4, IL7-4-1 and control (M82).
Values are presented as the mean ± SE of three to five biological replicates. Asterisks indicate significant differences from the control (M82) as assessed by Student's *t*-test ($P < 0.01$). *$P \leq 0.01$.

© 2020 The Authors.

O-sophoroside-glycoside, isorhamnetin-*3*-*O*-sophoroside-glycoside and kaempferol-*3*-*O*-sophoroside-glycoside given that these were distinct from those previously determined in the fruit.

Identification of seed-specific flavonoid UGTs in tomato

The genetic architecture of seed flavonoid metabolism has been elucidated in the model species Arabidopsis (Ishihara *et al*., 2016; Tohge *et al*., 2016) and in grass crop species, including rice and maize (Matsuda *et al*., 2015; Peng *et al*., 2017). Flavonoids, as with many specialized compounds, are modified by a wide range of specific reactions, including glycosylation, hydroxylation and methylation that results in generation of an enormous number of different structures (Alseekh *et al*., 2020). These compounds have additionally been documented to play vital roles in pathogen and predator defense as well as seed maturation and dormancy. However, to date the exact functions of the individual structures have rarely been studied. In tomato, the metabolic composition of the cultivated and wild species has been extensively investigated (Tieman *et al*., 2006a; Schilmiller *et al*., 2010; Tieman *et al*., 2012; Leong et al., 2019). Flavonoids are one of the largest plant metabolite groups and have known benefits both for the plants themselves and for human health. Flavonol glycosides are one of the major flavonoid classes present in *S. lycopersicum*. That said, the underlying mechanisms responsible for glycosylation and natural variation in the levels of these compounds are essentially unknown.

In this study, seed-specific flavonoids (quercetin-3-*O*-sophoroside-glycoside, isorhamnetin-3-*O*-sophoroside-glycoside and kaempferol-3-*O*-sophoroside-glycoside) were identified. Alterations in the levels of these metabolites mapped an mQTL for these metabolites to within the IL12-3/IL12-4 overlap region. The subsequent use of BILs (Ofner *et al*., 2016) and the gene expression profiles (Figures 3 and 4; Figure S2) allowed us to select a candidate gene underlying this variation. The same strategy was recently adopted to confirm the candidature of a lipase for the mQTL responsible for TAG and DAG breakdown and levels of fatty acid derived volatiles in tomato (Garbowicz *et al*., 2018). Similarly, the BILs have been used to identify the genes underlying acyl sugar variation and circadian clock in cultivated tomato (Müller *et al*., 2015; Fan *et al*., 2016). Further confirmation of the gene functions of *Solyc12g098600* and *Solyc12g096870* was provided by stable transgenic overexpression studies that resulted in marked increases (up to 13-fold) in quercetin-3-*O*-sophoroside-rhamnoside and other flavone derivatives (Figure S6) for *Solyc12g098600* and only one line showed four-fold increases in overexpressing *Solyc12g096870* (Figure S6). These results suggest that *Solyc12g098600* is the most likely gene responsible for the content of quercetin-3-*O*-sophoroside-rhamnoside. In addition, the levels of kaempferol-3-*O*-sophoroside-*O*-rhamnoside increased 2.5-fold in a line overexpressing *Solyc12g096870* (Figure S6). This suggests that regulating the change in the promoter is potentially linked to changes in expression as a putative mechanism for QTL.

Finally, results from transgenic plants with altered expression of *Solyc12g098600* and *Solyc12g096870* suggest that these enzymes catalyze the addition of a rhamnose moiety to quercetin diglycosides, and kaempferol diglycosides or isorhamnetin diglycosides, respectively (Figure 2b). In Arabidopsis, it was shown that the seed coat contains most of the quercetin-3-rhamnoside and flavonols of seed embryos (Kuhn *et al*., 2016; Muhlemann *et al*., 2018). Therefore, our data suggest that the modification of flavonols varies in a tissue-specific manner; therefore, we believe our study will help to understand the genetics of glycosylation and flavonol structures in tomato seeds better and may be of use in engendering strategies aimed at improving the nutritional aspects of plants. Furthermore, previous studies have shown that the accumulation of flavonoids such as quercetin and kaempferol derivatives tend to decrease during fruit ripening (Rohrmann *et al*., 2011). In contrast, our data showed that the quercetin-3-*O*-sophoroside-rhamnoside and kaempferol-3-*O*-sophoroside-rhamnoside highly accumulated in red ripe seeds, which may play a role in seed protection and dispersal in the *S. pennellii* tomato wild species.

Identification of glycoalkaloid UGTs in tomato seeds

Both enhanced nutritional quality and removal of antinutritional traits are needed to improve food. During crop domestication, antinutrient levels were reduced by selection and/or breeding; however, some of these substances remain in our food source. Use of wild germplasm, as a source of several important traits such as pathogen resistance, may therefore also be complicated by co-introduction of antinutritional compounds that, subsequently, need to be removed. SGA biosynthesis requires genes encoding UGTs, which add diverse sugar moieties to the steroidal alkaloid skeleton (McCue *et al*., 2005; Itkin *et al*., 2011). A comparative coexpression analysis between tomato and potato revealed the presence of operon-like clustering of the SGA biosynthesis genes (GAME) on chromosomes 7 and 12. In addition, Sonawane *et al*. (2016) proposed the main pathways of SGA metabolism in tomato following the profiling of eight wild tomato species accessions. Despite the identification of hundreds of SGAs, the enzymes responsible for their metabolism have only been reported in a couple of selected instances (Itkin *et al*., 2011, 2013; Schwahn *et al*., 2014). Our analysis revealed that the overlapping region of IL7-4 and IL7-4-1 harbors mQTL responsible for SGA changes in seed. Fine mapping using the BILs led to a narrowing of the QTL to a small genomic region harboring only 14 genes. These include four UGTs

© 2020 The Authors.
The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd,

in a cluster with a 2-oxoglutarate-dependent dioxygenase and a cytochrome P450, which were previously defined as a gene cluster that involved SGA biosynthesis in tomato and potato (Itkin *et al.*, 2011, 2013). Our results confirm the earlier findings using natural variation as an alternative experimental approach. To support our findings, we checked the expression of those genes (Bolger *et al.*, 2014) comparing the expression between the cultivated variety and the wild species. This study revealed a significantly different gene expression level between *S. lycopersicum* and *S. pennellii* in several tissues. Furthermore, sequence polymorphisms between the coding regions of *S. lycopersicum* and *S. pennellii* indicated that there are differences between both alleles. Interestingly, GAME 2 (*Solyc07g043410*) involved in the modification of glycoalkaloids showed three indels, each being 9-bp deletions, in the *S. lycopersicum* compared with the *S. pennellii*, which might change the reading frame and consequently produce different protein sequences or lead to premature termination. The higher gene expression of *Solyc07g043410* in the seeds of IL7-4-1 and IL7-4 compared with M82 (Figure S9) might support the degradation of mRNA in M82 and consequently the observed phenotype. This result implicated GAME 2 as being involved in the differences in glycoalkaloid levels observed in the seeds. However, further research will be required to confirm this hypothesis.

Comparing our seed mQTL observations with previously published fruit data (Alseekh *et al.*, 2015), we analyzed if there were conserved mQTL in the IL7-4 and IL7-4-1 in the fruit. Although there are 10 mQTL overlapping and are significantly conserved in two independent harvests. However, only one annotated glycoalkaloid derivative (1363.5 *m/z*) was lower in IL7-4 and IL7-4-1 (Alseekh *et al.*, 2015). Furthermore, another four mQTL for nitrogen-containing compounds, and another five mQTL for flavonoid derivatives and unknown compounds were reported in the fruit. Despite the fact that we have reported a glycoalkaloid mQTL in chromosome 7 in fruit data, our results here revealed a specific seed glycoalkaloid derivative at this location but, furthermore, were able to narrow down the mQTL to a genomic region with a relatively very small number of genes (14). Our results provide basic information that might be used for engendering approaches to increase beneficial and decrease undesirable nutrients in crops. Moreover, the result we reported here was not described in a recent genome-wide association study of an analysis in fruit (Zhu *et al.*, 2018). This provides an example of how populations from biparental crosses can have uses beyond that available in genome-wide association panels. Our study thus uncovered an important genomic region involved in glycoalkaloid production; however, as in the case of flavonoid variation described above, future work will be needed to uncover the hierarchy of control of their levels by the genes involved in this interval.

EXPERIMENTAL PROCEDURES

Plant materials and secondary metabolite measurement

Metabolite data were obtained from dry seeds of 76 ILs and their recurrent parent M82. Analysis was conducted as described by Tohge and Fernie (2010) using a high-performance LC system (Surveyor; Thermo Finnigan, USA), coupled to a Finnigan LTQ-XP system (Thermo Finnigan). In brief, the ILs and M82 were grown in pots containing artificial standard soil (70% white peat (v/v), 20% (v/v) vermiculite and 10% (v/v) sand under long-day conditions (250 μmol photons m^{-2} sec^{-1}, 16/8 h day/night cycle) at 22°C and 50% humidity, four seeds of each IL and the recurrent parent M82 were weighed, homogenized in liquid nitrogen and extracted with 80% methanol before LC-MS analysis in negative ion mode. Six to 10 seeds were cleaned and used per replicate (four biological replicates were used for each IL). All data were processed using XCALIBUR software version 2.1 (Thermo Fisher Scientific, http://www.thermofisher.com/). Metabolite identification and annotation were performed using standard compounds and tomato metabolomics databases (Moco *et al.*, 2006; Iijima *et al.*, 2008; Tohge *et al.*, 2009; Tohge and Fernie, 2010; Rohrmann *et al.*, 2011). Metabolome data are reported following the data reporting standards suggested by Fernie *et al.* (2011) and metadata is presented in Table S2. The same growth conditions were used to grow plants of stable transgenic lines and tissues of selected lines were collected and subjected to LC-MS as described above.

QTL mapping

Metabolites data were used for QTL mapping, each IL was compared (by *t*-test, at permissive $P \leq 0.05$ and strict $P \leq 0.01$ thresholds) with M82. If a trait value for an IL was significantly different from that for the reference genotype M82, the introgression was considered as harboring a QTL. The same approach was taken for the ILs as described previously (Alseekh *et al.*, 2015; Garbowicz *et al.*, 2018).

Heat maps

Heat maps were created by MEV software (http://mev.tm4.org), using the log2 of the metabolite fold-changes with respect to the recurrent parent and M82. Each cell in the heat map represents the $\log_2$ average fold-change value. Hierarchical clustering was applied to rows and columns using Pearson's correlation coefficient.

Fine mapping of the flavonol and glycoalkaloid mQTL

A backcross inbred line (BIL) population was constructed from a cross between the cultivated tomato *S. lycopersicum* (cv M82) and the wild species *S. pennellii* (LA0716) (Ofner *et al.*, 2016). The BILs were genotyped with the 10K SOLCAP SNP Chip analysis (Sim *et al.*, 2012). Polymorphic markers between the two parents were used for the mapping analysis. Seeds of BILs covering the targets regions of IL12-3 and IL12-4 for flavonoids mQTL and BILs cover IL7-4 and IL7-4-1 for glycoalkaloid mQTL were provided by Dani Zamir (Hebrew University of Jerusalem, Israel). Seeds from 15 and 10 lines covering IL12-3/IL12-4 and IL7-4/IL7-4-1, respectively. Seeds collected from plants were grown in the Western Galilee Experimental Station in Akko, Israel, in a completely randomized design with one plant per m^2. Seedlings were grown in greenhouses for 35–40 days and then transferred to the field. The field was irrigated with 320 m^3 of water per 1000 m^3 of field area throughout the season (Eshed and Zamir,

© 2020 The Authors.

1995). The generation of sub-ILs from IL12-3 is described in Garbowicz *et al.*, (2018), 16 sub-ILs were used to validate and narrow down the mQTL for flavonoid derivatives. Seeds were subjected to LC-MS analysis as described above for mapping the mQTL.

Cloning of UGTs and generation of stable transgenic overexpression lines

Leaf disk of tomato (*S. lycopersicum* "Moneymaker") plants were obtained from a stock culture initiated from seeds germinated *in vitro*. Transformation of *S. lycopersicum* was performed essentially as described by Tauberger *et al.* (2000). A pBINAR plasmid was prepared containing the nptII gene for kanamycin resistance and the *Solanum pennellii* alleles of *Solyc12g098600* (1392pb) and *Solyc12g096870* (1461pb) genes in the sense orientation under the control of a single constitutive (CaMV 35S) promoter, which was introduced into plants via an agrobacterium-mediated gene transfer procedure.

Phylogenetic analysis of UGT genes

Multiple sequence alignment of protein sequences was performed using CLUSTALW (http://www.genome.jp/tools/clustalw/). The phylogenetic tree was built using MEGA5.1 (Tamura *et al.*, 2011). The tree was inferred by using the neighbor-joining statistical method with 1000 bootstrap replicates.

Transient overexpression experiment

The full-length cDNA of *pene12g098600* and *pene12g096870* genes from *S. pennellii* was cloned into the pBI121 vector. Transient overexpression was performed in *Nicotiana benthamiana* leaves by the co-infiltration of an *Agrobacterium tumefaciens* strain containing PBI-*pene12g098600*, PBI-*pene12g096870* and the pBin61-p19 construct to overexpress a gene encoding the silencing suppressor p19 protein of tomato bushy stunt virus under the control of the CaMV 35S promoter (Zanor *et al.*, 2009). Leaves inoculated by agroinfiltrating of bacterial suspension (a mix of the *A. tumefaciens* strain containing PBI-12g098600, PBI-12g096870 and pBin61-p19 vectors in a 1:1:1 ratio). Agroinjected leaves were collected after 3 days, and subjected to LC-MS analysis as described above.

Gene expression analysis

Total RNA was extracted from different tissues as described by Bugos *et al.* (1995). First-strand cDNA synthesis prepared using 500 μg of total RNA and Superscript III reverse transcriptase (Invitrogen, http://www.invitrogen.com/), by mixing up to oligo (dT) 18 primer dNTPs (10 mM) and MilliQ-water, and incubated at 65°C for 5 min and chilled on ice. After Superscript III buffer, DTT (0.1 M) and Superscript III reverse transcriptase (200 U) were added. The mix was incubated 50 min at 50°C followed by 15 min at 70°C. Quantitative reverse transcription–PCR was used to evaluate the expression, using the fluorescent interacting dye SYBR Green (TaqMan) in an iCycler detection system (Bio-Rad, http://www.bio-rad.com/). The quantitative reverse transcription–PCR reactions were performed following the recommended thermal profile: 50°C for 2 min, 95°C for 10 min, followed by 40 cycles of 95°C for 15 sec and 60°C for 1 min. After 40 cycles, the specificity of the amplifications was tested by heating from 60 to 95°C at an increment of 1.9°C per min, resulting in melting curves. We used delta C_t (Schmittgen and Livak, 2008) and expression data were normalized to the reference gene elongation factor1a (Zanor *et al.*, 2009) (GenBank accession no. X14449) and ubiquitin3 (Osorio *et al.*, 2012) (GenBank accession no. X58253) (Table S4).

ACKNOWLEDGEMENTS

Work in Fernie Lab was supported by the Deutsche Forschungsgemeinschaft in the framework of Deutsche Israeli Project FE 552/12-1. In addition, SA and ARF acknowledge funding of the PlantaSYST project by the European Union's Horizon 2020 Research and Innovation Programme (SGA-CSA no. 664621 and no. 739582 under FPA no. 664620). SO acknowledgements her support provided by Ministerio de Ciencia e Innovación, University of Málaga (Spain) through the grant Ramon and Cajal program (RYC-09170) and Plan Propio. Work in the Last group was funded by US National Science Foundation grant IOS-PGRP-1546617. We thank the C. M. Rick Tomato Genetics Resource Center (TGRC) for providing the seeds for *S. pennellii* accessions.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no competing interests.

AUTHOR CONTRIBUTIONS

SA, TT and AF contributed to the conception and design of the experiments, ZL, RL, DZ, IO provide tomato seeds, SO and JV generated transgenic plants, SA preformed experiments and analysis, SA and AF wrote the manuscript.

DATA AVAILABILITY STATEMENT

All relevant data can be found within the manuscript and its supporting materials.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Genetic architecture of seed secondary metabolites QTL across the tomato introgression line populations.

Figure S2. Expression levels of *Solyc12g098600* and *Solyc12g096870* in seeds and other tissues.

Figure S3. Sequence comparisons of 1000 bp upstream coding region and *Solyc12g096870* and *Solyc12g098600*.

Figure S4. Alignment of protein sequences between *S. pennellii* and *S. lycopersicum* (cv M82).

Figure S5. Phylogenetic analysis of candidate UGT genes involved in flavonoid biosynthesis.

Figure S6. Functional validation of UGTs.

Figure S7. Levels of flavonol-3-*O*-triglycosides in transient overexpression experiment.

Figure S8. Levels of glycoalkaloid derivatives in leaves seeds in IL7-4, IL7-4-1, IL7-5 and M82.

Figure S9. Expression levels of *Solyc07g043410*, *Solyc07g043480*, *Solyc07g043490* and *Solyc07g043500* in seed of IL7-4, IL7-4-1 and M82.

Table S1. Secondary metabolites levels in the 76 *S. pennellii* introgression lines.

© 2020 The Authors.

Table S2. Reporting table for secondary metabolite measurements.

Table S3. Levels of quercetin-3-*O*-sophoroside-*O*-rhamnoside in dry seeds of *S. pennellii* wild-type accessions.

Table S4. Primers used for qPCR.

REFERENCES

Alseekh, S., Ofner, I., Pleban, T., Tripodi, P., Di Dato, F., Cammareri, M., Mohammad, A., Grandillo, S., Fernie, A.R. and Zamir, D. (2013) Resolution by recombination: breaking up *Solanum pennellii* introgressions. *Trends Plant Sci.* **18**, 536–538.

Alseekh, S., Perez de Souza, L., Benina, M. and Fernie, A.R. (2020) The style and substance of plant flavonoid decoration; towards defining both structure and function. *Phytochemistry*, **174**, 112347.

Alseekh, S., Tohge, T., Wendenberg, R. *et al.* (2015) Identification and mode of inheritance of quantitative trait loci for secondary metabolite abundance in tomato. *Plant Cell*, **27**, 485–512.

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.

Angelovici, R., Galili, G., Fernie, A.R. and Fait, A. (2010) Seed desiccation: a bridge between maturation and germination. *Trends Plant Sci.* **15**, 211–218.

Bentsink, L., Alonso-Blanco, C., Vreugdenhil, D., Tesnier, K., Groot, S.P. and Koornneef, M. (2000) Genetic analysis of seed-soluble oligosaccharides in relation to seed storability of Arabidopsis. *Plant Physiol.* **124**, 1595–1604.

Blankemeyer, J.T., White, J.B., Stringer, B.K. and Friedman, M. (1997) Effect of alpha-tomatine and tomatidine on membrane potential of frog embryos and active transport of ions in frog skin. *Food Chem. Toxicol.* **35**, 639–646.

Bolger, A., Scossa, F., Bolger, M.E. *et al.* (2014) The genome of the stress-tolerant wild tomato species *Solanum pennellii*. *Nat. Genet.* **46**, 1034–1038.

Bugos, R.C., Chiang, V.L., Zhang, X.H., Campbell, E.R., Podila, G.K. and Campbell, W.H. (1995) RNA isolation from plant tissues recalcitrant to extraction in guanidine. *Biotechniques*, **19**, 734–737.

Capanoglu, E., Beekwilder, J., Boyacioglu, D., Hall, R. and de Vos, R. (2008) Changes in antioxidant and metabolite profiles during production of tomato paste. *J. Agric. Food Chem.* **56**, 964–973.

Cataldi, T.R., Lelario, F. and Bufo, S.A. (2005) Analysis of tomato glycoalkaloids by liquid chromatography coupled with electrospray ionization tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **19**, 3103–3110.

Chandra, H.M. and Ramalingam, S. (2011) Antioxidant potentials of skin, pulp, and seed fractions of commercially important tomato cultivars. *Food Sci. Biotechnol.* **20**, 15–21.

Debeaujon, I., Nesi, N., Perez, P., Devic, M., Grandjean, O., Caboche, M. and Lepiniec, L. (2003) Proanthocyanidin-accumulating cells in *Arabidopsis testa*: regulation of differentiation and role in seed development. *Plant Cell*, **15**, 2514–2531.

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.

Eshed, Y. and Zamir, D. (1995) An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics*, **141**, 1147–1162.

Fan, P.X., Miller, A.M., Schilmiller, A.L., Liu, X.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. USA*, **113**, E239–E248.

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.

Ferreres, F., Taveira, M., Pereira, D.M., Valentao, P. and Andrade, P.B. (2010) Tomato (*Lycopersicon esculentum*) seeds: new flavonols and cytotoxic effect. *J. Agric. Food Chem.* **58**, 2854–2861.

Fraser, P.D., Enfissi, E.M., Goodfellow, M., Eguchi, T. and Bramley, P.M. (2007) Metabolite profiling of plant carotenoids using the matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Plant J.* **49**, 552–564.

Fridman, E., Carrari, F., Liu, Y.S., Fernie, A.R. and Zamir, D. (2004) Zooming in on a quantitative trait for tomato yield using interspecific introgressions. *Science*, **305**, 1786–1789.

Friedman, M. (2002) Tomato glycoalkaloids: role in the plant and in the diet. *J. Agric. Food Chem.* **50**, 5751–5780.

Friedman, M. and Levine, C.E. (1998) Dehydrotomatine content in tomatoes. *J. Agric. Food Chem.* **46**, 4571–4576.

Friedman, M., McQuistan, T., Hendricks, J.D., Pereira, C. and Bailey, G.S. (2007) Protective effect of dietary tomatine against dibenzo[a, l]pyrene (DBP)-induced liver and stomach tumors in rainbow trout. *Mol. Nutr. Food Res.* **51**, 1485–1491.

Fujiwara, Y., Kiyota, N., Hori, M. *et al.* (2007) Esculeogenin A, a new tomato sapogenol, ameliorates hyperlipidemia and atherosclerosis in ApoE-deficient mice by inhibiting ACAT. *Arterioscler. Thromb. Vasc. Biol.* **27**, 2400–2406.

Garbowicz, K., Liu, Z., Alseekh, S. *et al.* (2018) Quantitative trait loci analysis identifies a prominent gene involved in the production of fatty acid-derived flavor volatiles in tomato. *Mol. Plant*, **11**, 1147–1165.

Giovannucci, E. (1999) Tomatoes, tomato-based products, lycopene, and cancer: review of the epidemiologic literature. *J. Natl Cancer Inst.* **91**, 317–331.

Gomez-Romero, M., Segura-Carretero, A. and Fernandez-Gutierrez, A. (2010) Metabolite profiling and quantification of phenolic compounds in methanol extracts of tomato fruit. *Phytochemistry*, **71**, 1848–1864.

Iijima, Y., Nakamura, Y., Ogata, Y. *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. *et al.* (2013) Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. *Science*, **341**, 175–179.

Itkin, M., Rogachev, I., Alkan, N. *et al.* (2011) GLYCOALKALOID METABOLISM1 is required for steroidal alkaloid glycosylation and prevention of phytotoxicity in tomato. *Plant Cell*, **23**, 4507–4525.

Katsumata, A., Kimura, M., Saigo, H., Aburaya, K., Nakano, M., Ikeda, T., Fujiwara, Y. and Nagai, R. (2011) Changes in esculeoside A content in different regions of the tomato fruit during maturation and heat processing. *J. Agric. Food Chem.* **59**, 4104–4110.

Koornneef, M., Bentsink, L. and Hilhorst, H. (2002) Seed dormancy and germination. *Curr. Opin. Plant Biol.* **5**, 33–36.

Kozukue, N., Han, J.S., Lee, K.R. and Friedman, M. (2004) Dehydrotomatine and alpha-tomatine content in tomato fruits and vegetative plant tissues. *J. Agric. Food Chem.* **52**, 2079–2083.

Kozukue, N., Yoon, K.S., Byun, G.I., Misoo, S., Levin, C.E. and Friedman, M. (2008) Distribution of glycoalkaloids in potato tubers of 59 accessions of two wild and five cultivated *Solanum* species. *J. Agric. Food Chem.* **56**, 11920–11928.

Krause, M. and Galensa, R. (1992) Determination of naringenin and naringenin-chalcone in tomato skins by reversed phase HPLC after solid phase extraction. *Zeitschrift für Lebensmittel-Untersuchung und Forschung*, **194**, 29–32.

Kuhn, B.M., Errafi, S., Bucher, R., Dobrev, P., Geisler, M., Bigler, L., Zazimalova, E. and Ringli, C. (2016) 7-Rhamnosylated flavonols modulate homeostasis of the plant hormone auxin and affect plant development. *J. Biol. Chem.* **291**, 5385–5395.

Lazos, E.S., Tsaknis, J. and Lalas, S. (1998) Characteristics and composition of tomato seed oil. *Grasas Aceites*, **49**, 440–445.

Liu, Y.S., Gur, A., Ronen, G., Causse, M., Damidaux, R., Buret, M., Hirschberg, J. and Zamir, D. (2003) There is more to tomato fruit colour than candidate carotenoid genes. *Plant Biotechnol. J.* **1**, 195–207.

Leong, B.J., Lybrand, D.B., Lou, Y.R., Fan, P., Schilmiller, A.L. and Last, R.L. (2019) Evolution of metabolic novelty: a trichome-expressed invertase creates specialized metabolic diversity in wild tomato. *Sci. Adv.* **5**, eaaw3754.

Malecka, M. (2002) Antioxidant properties of the unsaponifiable matter isolated from tomato seeds, oat grains and wheat germ oil. *Food Chem.* **79**, 327–330.

© 2020 The Authors.
The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd,

Maloney, G.S., Kochevenko, A., Tieman, D.M., Tohge, T., Krieger, U., Zamir, D., Taylor, M.G., Fernie, A.R. and Klee, H.J. (2010) Characterization of the branched-chain amino acid aminotransferase enzyme family in tomato. *Plant Physiol.* **153**, 925–936.

Martin, C. and Li, J. (2017) Medicine is not health care, food is health care: plant metabolic engineering, diet and human health. *New Phytol.* **216**, 699–719.

Matsuda, F., Nakabayashi, R., Yang, Z., Okazaki, Y., Yonemaru, J., Ebana, K., Yano, M. and Saito, K. (2015) Metabolome-genome-wide association study dissects genetic architecture for generating natural variation in rice secondary metabolism. *Plant J.* **81**, 13–23.

McCue, K.F., Shepherd, L.V.T., Allen, P.V., Maccree, M.M., Rockhold, D.R., Corsini, D.L., Davies, H.V. and Belknap, W.R. (2005) Metabolic compensation of steroidal glycoalkaloid biosynthesis in transgenic potato tubers: using reverse genetics to confirm the in vivo enzyme function of a steroidal alkaloid galactosyltransferase. *Plant Sci.* **168**, 267–273.

Milner, S.E., Brunton, N.P., Jones, P.W., O'Brien, N.M., Collins, S.G. and Maguire, A.R. (2011) Bioactivities of glycoalkaloids and their aglycones from *Solanum* species. *J. Agric. Food Chem.* **59**, 3454–3484.

Moco, S., Bino, R.J., Vorst, O., Verhoeven, H.A., de Groot, J., van Beek, T.A., Vervoort, J. and de Vos, C.H. (2006) A liquid chromatography-mass spectrometry-based metabolome database for tomato. *Plant Physiol.* **141**, 1205–1218.

Muhlemann, J.K., Younts, T.L.B. and Muday, G.K. (2018) Flavonols control pollen tube growth and integrity by regulating ROS homeostasis during high-temperature stress. *Proc. Natl. Acad. Sci. USA*, **115**, E11188–E11197.

Muir, S.R., Collins, G.J., Robinson, S., Hughes, S., Bovy, A., Ric De Vos, C.H., van Tunen, A.J. and Verhoeyen, M.E. (2001) Overexpression of petunia chalcone isomerase in tomato results in fruit containing increased levels of flavonols. *Nat. Biotechnol.* **19**, 470–474.

Müller, N.A., Wijnen, C.L., Srinivasan, A. *et al.* (2015) Domestication selected for deceleration of the circadian clock in cultivated tomato. *Nat. Genet.* **48**, 89.

Nunes-Nesi, A., Alseekh, S., de Oliveira Silva, F.M. *et al.* (2019) Identification and characterization of metabolite quantitative trait loci in tomato leaves and comparison with those reported for fruits and seeds. *Metabolomics*, **15**, 46.

Ofner, I., Lashbrooke, J., Pleban, T., Aharoni, A. and Zamir, D. (2016) *Solanum pennellii* backcross inbred lines (BILs) link small genomic bins with tomato traits. *Plant J.* **87**, 151–160.

Osorio, S., Alba, R., Nikoloski, Z., Kochevenko, A., Fernie, A.R. and Giovannoni, J.J. (2012) Integrative comparative analyses of transcript and metabolite profiles from pepper and tomato ripening and development stages uncovers species-specific patterns of network regulatory behavior. *Plant Physiol.* **159**, 1713–1729.

Peng, M., Shahzad, R., Gul, A. *et al.* (2017) Differentially evolved glucosyltransferases determine natural variation of rice flavone accumulation and UV-tolerance. *Nat. Commun.* **8**, 1975.

Quadrana, L., Almeida, J., Asis, R. *et al.* (2014) Natural occurring epialleles determine vitamin E accumulation in tomato fruits. *Nat. Commun.* **5**, 4027.

Rohrmann, J., Tohge, T., Alba, R. *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.

Rousseaux, M.C., Jones, C.M., Adams, D., Chetelat, R., Bennett, A. and Powell, A. (2005) QTL analysis of fruit antioxidants in tomato using *Lycopersicon pennellii* introgression lines. *Theor. Appl. Genet.* **111**, 1396–1408.

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.

Ruan, Y.L., Patrick, J.W., Bouzayen, M., Osorio, S. and Fernie, A.R. (2012) Molecular regulation of seed and fruit set. *Trends Plant Sci.* **17**, 656–665.

Sawai, S., Ohyama, K., Yasumoto, S. *et al.* (2014) Sterol side chain reductase 2 is a key enzyme in the biosynthesis of cholesterol, the common precursor of toxic steroidal glycoalkaloids in potato. *Plant Cell*, **26**, 3763–3774.

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. *et al.* (2006) Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nat. Biotechnol.* **24**, 447–454.

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.

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.

Schilmiller, A.L., Charbonneau, A.L. and Last, R.L. (2012) Identification of a BAHD acetyltransferase that produces protective acyl sugars in tomato trichomes. *Proc. Natl Acad. Sci. USA*, **109**, 16377–16382.

Schmittgen, T.D. and Livak, K.J. (2008) Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.* **3**, 1101–1108.

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.

Sim, S.C., Durstewitz, G., Plieske, J. *et al.* (2012) Development of a large SNP genotyping array and generation of high-density genetic maps in tomato. *PLoS One*, **7**(7), e40563.

Sonawane, P.D., Pollier, J., Panda, S. *et al.* (2016) Plant cholesterol biosynthetic pathway overlaps with phytosterol metabolism. *Nat. Plants*, **3**, 16205.

Stevens, R., Buret, M., Duffé, 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.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739.

Tauberger, E., Fernie, A.R., Emermann, M., Renz, A., Kossmann, J., Willmitzer, L. and Tretheway, R.N. (2000) Antisense inhibition of plastidial phosphoglucomutase provides compelling evidence that potato tuber amyloplasts import carbon from the cytosol in the form of glucose-6-phosphate. *Plant J.* **23**, 43–53.

Tieman, D., Bliss, P., McIntyre, L.M. *et al.* (2012) The chemical interactions underlying tomato flavor preferences. *Curr. Biol.* **22**, 1035–1039.

Tieman, D., Taylor, M., Schauer, N., Fernie, A.R., Hanson, A.D. and Klee, H.J. (2006a), Tomato aromatic amino acid decarboxylases participate in synthesis of the flavor volatiles 2-phenylethanol and 2-phenylacetaldehyde. *Proc. Natl. Acad. Sci. USA*, **103**(21), 8287–8292.

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

Tohge, T., de Souza, L.P. and Fernie, A.R. (2014) Genome-enabled plant metabolomics. *J. Chromatogr. B*, **966**, 7–20.

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., Nishiyama, Y., Hirai, M.Y. *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.

Tohge, T., Scossa, F., Wendenburg, R. *et al.* (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.

Tohge, T., Wendenburg, R., Ishihara, H. *et al.* (2016) Characterization of a recently evolved flavonol-phenylacyltransferase gene provides signatures of natural light selection in Brassicaceae. *Nat. Commun.* **7**, 12399.

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

Torres, C.A., Davies, N.M., Yanez, J.A. and Andrews, P.K. (2005) Disposition of selected flavonoids in fruit tissues of various tomato (*Lycopersicon esculentum* mill.) Genotypes. *J. Agric. Food Chem.* **53**, 9536–9543.

Toubiana, D., Batushansky, A., Tzfadia, O., Scossa, F., Khan, A., Barak, S., Zamir, D., Fernie, A.R., Nikoloski, Z. and Fait, A. (2015) Combined correlation-based network and mQTL analyses efficiently identified loci for branched-chain amino acid, serine to threonine, and proline metabolism in tomato seeds. *Plant J.* **81**, 121–133.

© 2020 The Authors.
The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd,

Toubiana, D., Semel, Y., Tohge, T., Beleggia, R., Cattivelli, L., Rosental, L., Nikoloski, Z., Zamir, D., Fernie, A.R. and Fait, A. (2012) Metabolic profiling of a mapping population exposes new insights in the regulation of seed metabolism and seed, fruit, and plant relations. *PLoS Genet.* **8**, e1002612.

Valdez-Morales, M., Espinosa-Alonso, L.G., Espinoza-Torres, L.C., Delgado-Vargas, F. and Medina-Godoy, S. (2014) Phenolic content and antioxidant and antimutagenic activities in tomato peel, seeds, and byproducts. *J. Agric. Food Chem.* **62**, 5281–5289.

Willcox, J.K., Catignani, G.L. and Lazarus, S. (2003) Tomatoes and cardiovascular health. *Crit. Rev. Food Sci. Nutr.* **43**, 1–18.

Zanor, M.I., Osorio, S., Nunes-Nesi, A. *et al.* (2009) RNA interference of LIN5 in tomato confirms its role in controlling Brix content, uncovers the influence of sugars on the levels of fruit hormones, and demonstrates the importance of sucrose cleavage for normal fruit development and fertility. *Plant Physiol.* **150**, 1204–1218.

Zhang, Y., Butelli, E., Alseekh, S. *et al.* (2015) Multi-level engineering facilitates the production of phenylpropanoid compounds in tomato. *Nat. Commun.* **6**, 8635.

Zhu, G.T., Wang, S.C., Huang, Z.J. *et al.* (2018) Rewiring of the fruit metabolome in tomato breeding. *Cell*, **172**, 249–261.e12.

© 2020 The Authors.
The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd,