Received Date: 19-Mar-2015
Revised Date: 15-Jun-2015
Accepted Date: 16-Jun-2015
Article Type: Original Article
Running title: ***Anthocyanin biosynthesis in Solanaceous species***

Ectopic expression of snapdragon transcription factors facilitates the identification of genes encoding enzymes of anthocyanin decoration in tomato

Takayuki Tohge[1], Yang Zhang[3], Silke Peterek[2], Andrea Matros[2], Ghanasyam Rallapalli[4], Yudelsy A. Tandrón[2], Eugenio Butelli[3], Kalyani Kallam[3], Norbert Hertkorn[5], Hans-Peter Mock[2], Cathie Martin[3], Alisdair R. Fernie[1*]

[1] Max-Planck-Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam-Golm, Germany
[2] Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstr. 3, D-06466 Gatersleben, Germany
[3] John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UA, United Kingdom
[4] The Sainsbury Laboratory, Norwich Research Park, Colney, Norwich, UK NR4 7UH
[5] Helmholtz Zentrum München, German Research Center for Environment and Health, GmbH, Institute of Ecological Chemistry, Ingolstaedter Landstraße 1, D-85764 Neuherberg, Germany

Corresponding Author:
Alisdair R. Fernie
Max-Planck-Institute of Molecular Plant Physiology,
Am Muehlenberg 1,
14476 Potsdam,
Germany
E-Mail: fernie@mpimp-golm.mpg.de
*Corresponding author: Alisdair R. Fernie
Address: Max-Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1

14476, Potsdam-Golm, Germany
Tel: +49 33 1567 8211
Fax: +49 33 1567 8408

Keywords: anthocyanin metabolism, tomato, Solanaceous species, LC/MS profiling, GC/MS profiling, RNAseq, acyltransferase, glycosyltransferase.

ABSTRACT

Given the potential health benefits of polyphenolic compounds in the diet, there is a growing interest in the generation of food crops enriched with health-protective flavonoids. We undertook a series of metabolite analyses of tomatoes ectopically expressing the *Delila* and *Rosea1* transcription factor genes from snapdragon, paying particular attention to changes in phenylpropanoids compared to controls. These analyses revealed multiple changes including depletion of rutin and naringenin-chalcone, enhanced levels of anthocyanins and phenylacylated flavonol derivatives. We isolated and characterised the chemical structures of the two most abundant anthocyanins which were shown by nuclear magnetic resonance spectroscopy to be delphinidin-3-(4'''-*O*-*trans*-*p*-coumaroyl)-rutinoside-5-*O*-glucoside and petunidin-3-(4'''-*O*-*trans*-*p*-coumaroyl)-rutinoside-5-*O*-glucoside. By performing RNA sequencing on both the purple fruit and wild type fruit we collected important information concerning the relative expression of both structural and transcription factor genes. Integrative analysis of the transcript and metabolite datasets provided compelling evidence of the nature of all genes involved in anthocyanin biosynthesis including those encoding species-specific anthocyanin decoration enzymes. One gene, SlFdAT1 (Solyc12g088170), predicted to encode a flavonoid-3-*O*-rutinoside-4'''-phenylacyltransferase was characterized by assays of recombinant protein and overexpression assays in tobacco. The combined data are discussed in the context of both our current understanding of phenylpropanoid metabolism in Solanaceous species, and evolution of flavonoid decorating enzymes and their transcriptional networks in different plant species.

INTRODUCTION

The relationship between nutrition and prevention of human diseases has been the topic of several recent reviews (Fitzpatrick et al., 2012; Martin, 2013; Martin et al., 2013; Schwahn et al., 2014). Phenolics, a widespread group of secondary plant metabolites, have multiple functions in plants, including UV-B protection (Kusano et al., 2011), the control of growth and developmental processes (Vanholme et al., 2012) as well as defence against herbivores and pathogens (Brechenmacher et al., 2010; Huang et al., 2010). Their widespread distribution amongst seed plants (Tohge et al., 2013, 2013), and their ability to scavenge free radicals and reduce oxidative damage (Gutteridge, 1994; Halliwell, 2012), means that flavonoids and related phenolics have been identified as important bioactive molecules in the human diet. The beneficial influence of phenolics on a number of human diseases has been reported including the prevention of cancer (Hollman et al., 1996; Le Marchand, 2002), dementia (Commenges et al., 2000), atherosclerosis (Aviram and Fuhrman, 2002) and coronary heart disease (Hollman et al., 1996; Mojzisova and Kuchta, 2001). Consequently, there is an increasing interest in developing alternative food sources which are rich in phenolic compounds. The broad spectrum of biological properties of individual phenolics, their bioavailability and the range of levels in different foodstuffs are generally uncharacterised and even the chemical identity of the individual compounds themselves has often not been established. Nevertheless, elucidation of these compounds and their biological properties provide an important foundation upon which strategies for metabolic engineering or biofortification can be designed.

Given both the volume consumed and the variation in form of tomato-based products (such as salad, puree, pasta sauce and ketchup), tomato (*Solanum lycopersicum)* is one of the most important vegetables in the human diet (worldwide) and therefore can be regarded an ideal vehicle for enhancing phenylpropanoid intake. However, tomato cultivars contain relatively low levels of phenolic compounds (Le Gall et al., 2003; Willits et al., 2005) and anthocyanins are normally absent in tomato fruit (Jones et al., 2003). Several different strategies to modify the biosynthesis of phenolics and hence alter the composition of tomato fruit flavonoids have been tested. The most common approach being the altered expression of specific endogenous genes including *ANT*1, encoding a Myb transcription factor, *AFT* encoding a Myb

transcription factor and *DE-ETIOLATED* 1, encoding a chromatin remodelling factor (Jones et al., 2003; Mathews et al., 2003; Enfissi et al., 2010) and the ectopic expression of petunia chalcone isomerase or maize transcription factors *LC* and *C1* (Muir et al., 2001; A. et al., 2002). Whilst increased levels of flavonoids were reported in all instances, in none of these studies were anthocyanins detected at appreciable levels throughout the ripe tomato fruit, their accumulation being confined to the surface layers.

Ectopic fruit-specific expression of the snapdragon transcription factors Delila (*Del*, bHLH) and Rosea1 (*Ros1*, Myb), however, resulted in accumulation of anthocyanins throughout the fruit, to substantially higher levels than previously achieved by metabolic engineering strategies in tomato (Butelli et al., 2008). The control of anthocyanin biosynthesis normally involves a complex of Myb, bHLH and WDR proteins called the MBW complex; (Ramsay and Glover, 2005). Experiments involving over-expression of a Myb transcription factor only are dependent on endogenous WDR and bHLH proteins for anthocyanin production. The WDR protein is normally constitutively expressed in plant cells, so it was the combined expression of the bHLH and the Myb transcription factors that overcame the limited bHLH expression in tomato fruit and resulted in induction of *F3'5'H* and high levels of anthocyanin accumulation throughout *Del/Ros1* tomato fruit (Bovy et al., 2002; Butelli et al., 2008).

As evidenced by this example (and those described above), the function of transcription factors is often studied by overexpression since, given the size of transcription factor families, the consequences of down-regulating individual transcription factors are often masked by their redundancy of function. Considerable advances have been made in functional characterization of numerous members of bHLH and Myb classes of transcription factor following this approach (Borevitz et al., 2000; Grotewold et al., 2000; Hirai et al., 2007; Sonderby et al., 2007; Dal Cin et al., 2011; Kong et al., 2012). Analysis of anthocyanins has been greatly improved following the proof-of-concept study that took this approach – namely the combined transcriptomic and metabolomic evaluation of the AtMYB75 (AtPAP1), activation-tagged Arabidopsis line (Tohge et al., 2005) and follow-up research based on this study for functional characterization of genes encoding anthocyanin

glycosyltransferases and acyltransferases (Luo et al., 2009; Yonekura-Sakakibara et al., 2012). Despite the growing use of guilt-by-association approaches to correlate changes in transcripts/ transcription factors and metabolite levels in tomato (see Carrari et al., 2006; Mounet et al., 2009; Rohrmann et al., 2011), more comprehensive analysis of chemical changes resulting from targeted modification of transcription factor activity in crop species are largely lacking.

Here we describe comprehensive analysis of the phenolic content of *Del/Ros1* purple tomatoes in comparison to the wild type control as well as performing primary metabolic profiling and RNA sequencing for comparing fruit of the transgenic line and the wild type. We identified considerable alterations in a total of 113 compounds which included seven anthocyanins and 18 flavonol derivatives. We characterised the major anthocyanins as delphinidin-3-*O*-[4'''-(*trans*-*p*-coumaroyl)-*α*-L-rhamnopyranosyl-(1->6)-*β*-D-glucopyranoside]-5-*O*-*β*-D-glucopyranoside (TA1, commonly referred to as nasunin and violanin) and petunidin-3-*O*-[4'''-(*trans*-*p*-coumaroyl)-*α*-L-rhamnopyranosyl-(1->6)-*β*-D-glucopyranoside]-5-*O*-*β*-D-glucopyranoside (TA2, petanin) which have already been characterized in eggplant and petunia, respectively. Integration of metabolomic and transcriptomic data allowed us to speculate about the genetic and biochemical mechanisms underlying the alterations in these profiles. Furthermore, phylogenetic analysis of candidates for target genes of *Del/Ros1*, identified by the integrated approach, suggested functions for anthocyanin decorating enzymes, namely Solyc10g083440 for anthocyanin-3-*O*-glucosyltransferase (SlA3GlcT), Solyc12g098590 for anthocyanin-5-*O*-glucosyltransferase (SlA5GlcT), Solyc09g059170 for anthocyanin-3-*O*-glucoside-6''-*O*-rhamnosyltransferase (SlA3Glc6''RhaT), and Solyc12g088170 for anthocyanin-3-*O*-rutinoside-4'''-*O*-phenylacyltransferase. Further investigation of the function of Solyc12g088170 (SlFdAT1) using recombinant enzyme assays and metabolite profiling of transgenic tobacco confirmed that *SlFdAT1* encodes a flavonoid-3-*O*-rutinoside-4'''-*O*-phenylacyltransferase (SlFd3Glc6''Rha4'''PAT). Cross-species comparisons of anthocyanin acyltransferases suggested that flavonoid-3-*O*-rutinoside-4'''-*O*-phenylacyltransferases are well conserved in Solanaceous species but evolved after the family diverged from its closest relatives.

RESULTS

LC-MS-based metabolite profile of *Del*/*Ros1* transgenic tomatoes

We undertook LC-MS based secondary metabolite profiling of extracts from whole fruits, flesh and peel from wild type and the *Del*/*Ros1* purple tomatoes harvested at Breaker + 1week (B+1w)(Butelli et al., 2008). Representative chromatograms are shown in Figure 1A and Supplemental Figure S1, in which it can clearly be seen that the *Del*/*Ros1* line has a richer diversity of secondary metabolites than the control in both whole fruit and peel samples. Flavonoid (see aglycone chemical structures of Figure 1B) across both genotypes a total of seven anthocyanins, four flavonols, ten phenylacylated flavonols and four naringenin derivatives were identified and annotated. The anthocyanins and phenylacylated-flavonols were essentially detected only in the *Del*/*Ros1* fruit, whilst large quantitative differences were seen in the flavonols and naringenin-derivatives between the lines (Figure 1C, Supplemental Table 1). Indeed, focusing on the peel and whole fruit samples; phenylacylated-flavonols were 4-200 times lower in abundance in MicroTom fruit than in *Del*/*Ros1* tomatoes. By contrast, naringenin-chalcone (NrC) levels in the *Del*/*Ros1* fruit were only 60% those observed in the WT control, although naringenin derivatives were 1.3-6.0-fold higher than in the WT control and flavonol contents were invariant. Considering other pathways of secondary metabolism, as shown on the metabolic map of Supplemental Figure 2, Tomatine-related glycoalkaloids were realtively 2-fold higher in *Del*/*Ros1*, whilst the esculeoside type glycoalkaloids were slightly lower (Supplemental Figure S2 and Supplemental Table S1). Given that reduction of tomatine-type glycoalkaloid is a strong indicator of fruit ripening (Schwahn et al., 2014; Tohge et al., 2014), the latter observations of relatively higher level of tomatine-type glycoalkaloid in *Del/Ros1* likely reflect a delay of over-ripening, confirming earlier observations (Zhang et al., 2013).

GC-MS-based metabolite profile of *Del*/*Ros1* purple tomato

A total of 67 metabolites including amino acids and polyamines, organic acids, cell wall precursors, sugars, sugar phosphates, chlorogenic acids and nucleobases, were detected using our GC-MS platform. The whole purple tomato fruit were characterized by slight decreases in ornithine and tryptophan and major decreases in phenylalanine, alanine, a putative galactose peak and threitol compared to control fruit. In contrast, major increases were observed in quinate, a putative galacturonate

and, most dramatically, in four CGAs (Table 1 and Supplemental Figure S2). When the tissues were dissected metabolic changes were far more prominent. Thirty compounds (eight which increased and 22 which decreased), were altered in abundance in fruit flesh. Quinate, 2OG, a putative galacturonate and all isomers of CGA were increased significantly in flesh of purple tomatoes compared to controls. Thirteen amino acids and benzoate, nicotinate, galactose and maltose peaks, adenine, threitol and urea were lower in flesh of purple fruit than in controls. The picture for purple skin was similar with 25 metabolites altered in abundance in the transgenic line (15 which increased and 10 which decreased) compared to controls. In this tissue, lysine, methionine, putrescine, and threonate increased whilst alanine, beta-alanine, isoleucine and tyrosine decreased. In other classes of metabolites there were increases in 2OG and malate but decreases in nicotinate, a putative citramalate, galactose peaks as well as sucrose, glycerol and threitol in purple peel.

Comparative gene expression analysis in *Del/Ros1* and wild type tomatoes

To complement these metabolomic analyses we undertook RNA-seq analysis on extracts from whole fruit harvested at Breaker + 1week (B+1w) and Breaker + 4weeks (B+4w) developmental stages (Supplemental Table S2). Changes in gene expression were evaluated on a global scale using the most recent update of the tomato MapMan mapping files (Urbanczyk-Wochniak et al., 2006) (http://mapman.gabipd.org/web/guest)(Supplemental Figure S3). Genes associated with flavonoid, phenylpropanoid, cell wall, wax and aromatic amino acid biosynthesis as well as lipid metabolism and histidine degradation were upregulated in purple tomatoes whilst genes involved in protein and branched chain amino acid synthesis, serine, glycine and cysteine synthesis, cell wall modification, minor cell wall modifications and hormone metabolism as well as redox dismutases and catalases were downregulated in both developmental stages of purple tomatoes compared to controls. In addition, there were some developmental stage-specific differences between the genotypes. At the B+1w stage transcripts associated with the Calvin-Benson cycle and photosystems were up-regulated, as well as the TCA cycle (which is consistent with the elevated levels of 2OG and malate), sugar metabolism and sulphur assimilation in the purple fruit whilst transcripts associated with Gln, Pro and Arg synthesis relatively decreased, as did those for the branched chain amino acids (consistent with the observed changes in the levels of the metabolites themselves)

and most of lipid metabolism was downregulated. In summary, these changes in transcripts were consistent with a delay in ripening as previously observed (Zhang et al., 2013). At the B+4w stage the only specific differences were the downregulation of transcripts associated with the Calvin-Benson cycle, the photosystems and the TCA cycle.

We next performed a targeted analysis of genes associated with anthocyanin metabolism. Figure 2A presents the gene expression of known or previously annotated genes which are involved in phenylalanine, phenylpropanoid and flavonoid biosynthesis in tomato. By this approach we were able to annotate a total of 37 genes comprising five genes of phenylalanine biosynthesis (CM, chorismate mutase; PDT, prephenate aminotransferase; ADT, arogenate dehydratase), 13 genes of general phenylpropanoid metabolism (PAL, phenylalanine ammonia-lyase; C4H, cinnamate-4-hydroxylase; 4CL, 4-coumarate CoA ligase), three genes of hydroxycinnamate biosynthesis (HCT, hydroxycinnamoyl-Coenzyme A shikimate/quinate hydroxycinnamoyltransferase; HQT, CoA:quinate hydroxycinnamoyl transferase) and 16 flavonoid biosynthetic genes (CHS, chalcone synthase; CHI, chalcone isomerase; FOMT, flavonol-*O*-methyltransferase; F3H, flavanone-3-hydroxylase; F3'H, flavonoid-3′-hydroxylase; F3'5'H, flavonoid-3′5′-hydroxylase; FLS, flavonol synthase; DFR, dihydroflavonol reductase; ANS, anthocyanidin synthase; AGT, flavonoid-3-*O*-glycosyltransferase)(Figure 2B).

All genes of flavonoid-anthocyanin biosynthesis were upregulated (5-4088 and 5-1820-fold-change in B+1w and B+4w respectively) with the exception of the F3'H and FLS involved in flavonol specific biosynthetic branch, and FOMTs which are involved in steps of quercetin and cyanidin production and in *O*-methylation of trichome-specific flavonol aglycones. Furthermore, the genes which are not involved in anthocyanin biosynthesis, namely hydroxycinnamate biosynthetic gene, HQT were not upregulated.

We also analysed individual genes upregulated at different developmental stages. Table 2 presents the upregulated genes in *Del*/*Ros1* in both B+1 and B+4 stages of fruit ripening. Gene counts were assessed as described Rallapalli *et al*., (2014) and varied more than 8-fold in purple fruit compared to WT. Twelve transcripts,

(nine flavonoid biosynthetic genes and three miscellaneous genes), namely, *F3'5'H, ANS, DFR, GST*, *FFT-like, AnthOMT, bHLH, PDF*, *ACLB-2,* UGTs and an AAT were more than 100-fold upregulated in *Del*/*Ros1.* The upregulated genes included two additional flavonoid glycosyltransferase genes (Solyc09g059170, Solyc12g098590), one gene associated with the shikimate pathway and four additional genes associated with phenylpropanoid metabolism (Table 2).

We compared the upregulated genes with the genes reported to be upregulated in the *Ant1* overexpression mutant which showed intense purple pigmentation in seedlings (Mathews et al., 2003)(Table 2). Based on ITAG2.3 (International Tomato Annotaiton Group gene identifier) Sol annotation, EST contigs were searched using BLASTN and converted to Solyc IDs. Of the up-regulated genes in the *Del*/*Ros1* line, CHI, a putative A3GT, a putative A5GT, GST and FFT were also upregulated in the *Ant1*-overexpression experiments. We further interrogated this dataset by looking for conserved changes between the two ripening stages.

We compared genes upregulated in purple tomatoes relative to control fruit to those upregulated in Arabidopsis plants constitutively expressing the transcription factor PAP1 (*pap1-D*; Tohge et al., 2005). Venn diagrams of upregulated genes are shown in Figure 2B with a total of 374 genes upregulated more than 8-fold in *Del/Ros1* at B+1w and 36 at B+4w. Twenty nine genes were upregulated more than 8-fold compared to WT, at both stages. We next compared these conserved genes to their Arabidopsis homologous which were upregulated in the *pap1*-D mutant – and nine of the 29 are common (Table 2). There was considerably less overlap for down-regulated genes with only 28 genes being more than 8-fold downregulated in *Del/Ros1* at B+1w only, whereas 49 were downregulated at B+4 only, and four were downregulated at both time points. The same comparison was made to the *pap1-D* mutant, but none of the transcripts were commonly downregulated. These analyses highlight both conserved and species-specific differences and likely relate to differences in anthocyanin decoration between tissues and species analyzed as well as indirect transcriptional responses to the overexpression of the transcription factors.

Characterization of anthocyanins in tomato fruit

To assess pathway structure and regulation more fully, detailed structural data of the constituent metabolites were necessary. The six most abundant anthocyanins in tomato were separated and characterised using fruit material from the purple tomato line. The crude extract was processed by preparative HPLC and the peaks, corresponding to the main compounds detectable at 535 nm, were collected (Supplemental Figure S4 and Supplemental file S1). Pure fractions of the main anthocyanins were then analysed by LC/MS with and without acid hydrolysis to confirm the aglycone structure and sugar moieties (Supplemental Figure S5 and Supplemental file S1). The molecular ions $[M]^+$ at 919 and 933 with the annotated molecular formula $C_{42}H_{47}O_{23}$ and $C_{43}H_{49}O_{23}$, respectively).

The structures of the two main anthocyanins in tomato (TA1 and TA2) were elucidated using nuclear magnetic resonance (NMR; Supplemental file S2). In the ^{1}H NMR spectrum of TA1, the most significant feature was the absence of a methoxy group singlet at δ 4.00 ppm (present in compound V), attached to the C-3' position. The NMR spectra also confirmed the presence of two glucose molecules and one rhamnose molecule and one molecule of *p*-coumaric acid in each anthocyanin. In the *p*-coumaric acid moiety, 2''- and 3''-protons had a large coupling constant (J = 15.9 *Hz*) for both of the compounds. Therefore, the olefinic part of the *p*-coumaric acid moiety was concluded to exhibit a *trans*-configuration. In the rhamnose, the triplet signal for H-4 appeared to be shifted downfield, thus, the hydroxyl group at position 4 was acylated with *p*-coumaric acid in TA1 and TA2. This finding was confirmed by the occurrence of a cross peak in the heteronuclear multiple bond correlation spectrum between H-4 (rhamnose) and the carbonylic carbon (*p*-coumaric acid) [4.91/169.20 ppm]. In both of these compounds, the doublet signals of their glucose anomeric protons appeared at δ ca. 5.5 and ca. 5.2 ppm, respectively (glucose A and B) with a J = ca. 7.8 *Hz*, indicating a *β*-D-glucopyranose form. Moreover, the anomeric proton signal in the rhamnose moiety appeared as a singlet at δ 4.71 ppm and the methyl group appeared as a doublet (δ ca. 1.0; J = 6.2 Hz) suggesting the presence of the *α*-L-rhamnopyranose form. Finally, by analysis of nuclear Overhauser and exchange spectroscopy and heteronuclear multiple bond correlation spectra, the glucose A and B residues were found to be attached to the OH-3 and OH-5 respectively of the corresponding anthocyanidin through glycosidic bonds. These findings showed the

proposed structures of TA1 and TA2 to be delphinidin-3-O-[4'''-(*trans-p*-coumaroyl)-*α*-L-rhamnopyranosyl-(1->6)-*β*-D-glucoside]-5-O-*β*-D-glucoside (commonly referred to as nasunin) and petunidin-3-*O*-[4'''-(*trans-p*-coumaroyl)-*α*-L-rhamnosyl-(1->6)-*β*-D-glucoside]-5-*O*-*β*-D-glucoside (commonly referred to as petanin), respectively (Figure 3A).

Presence of anthocyanin-3-*O*-(4'''-*O*-phenylacyl)-rutinoside-5-*O*-glucosides in the plant kingdom

Having established that the ectopic expression of snapdragon transcription factors in tomato fruits can led to the synthesis of nasunin and petanin pigments, we next analyzed the anthocyanins present in tomato leaves of the common processing tomato cultivar M82. As shown in Figure 3B, M82 young leaf contains the same anthocyanins which were observed in *Del/Ros1* tomato fruits. Furthermore, the wild species, *Solanum pennellii*, which is characterized by its extreme stress tolerance (Bolger et al., 2014), also accumulates both of these compounds as major anthocyanins in leaves (Figure 3B). Furthermore, searching recently published RNA-seq data revealed that anthocyanin biosynthetic genes are relatively much higher expressed in seedlings and mature leaves than roots of both *S. lycopersicum* and *S. pennellii* (Supplemental Figure S6). Thus the expression levels are in keeping with the metabolite contents we observed. Our results suggested that there was conservation of anthocyanin decoration, between different tomato species so we searched the KNApSAcK database (http://kanaya.naist.jp/KNApSAcK/, (Afendi et al., 2012) and literature cited in the "Handbook of Natural Flavonoids" edited by Harborne and Baxter (2010), to assess the distribution of these specific flavonoids further. We searched both for all possible 4'''-*O*-phenylacylated anthocyanin-3-*O*-rutinoside-5-*O*-glucoside and for phenylacylated flavonol-3-*O*-rustinosides as summarized in Table 3. Amongst the decorated anthocyanins listed in the databases, 37 phenylacylated delphinidin derivatives, four contain the -*O*-(4'''-phenylacyl-rutinoside)-5-*O*-glucoside moiety, the same as TA1 and TA2. When anthocyanin-3-*O*-(4'''-phenylacyl-rutinoside)-5-*O*-glucosides derivatives only were selected, five petunidin, seven malvidin, two peonidin, two pelargonidin and a cyanidin derivative were identified belonging to this class. Interestingly, the major anthocyanins detected in this study in tomato, namely nasunin and petanin, have both been identified previously in other Solanaceous species. Nasunin (also named as Violanin) is particularly well

characterized in Asian eggplant (*S. melongena*; Fraser and Chapple, 2011; Lepelley et al., 2012)), as well as potato (Alseekh et al., 2015), pepper (Butelli et al., 2012) and petunia (Luo et al., 2008) although it is also present in *Iris tingitana* (Harborne, 1964; Nerdal and Andersen, 1992) and *Viola tricolor* (Handa and Mattoo, 2010), whereas petanin is particularly well characterized in petunia (Luo et al., 2008), as has been described in potato, *Solanum nigrum* and Iris (Handa and Mattoo, 2010). Other anthocyanin derivatives exhibiting 3-*O*-(4'''-phenylacyl-rutinoside)-5-*O*-glucoside decorations have mostly been described in different Solanaceous species (Petunia species, eggplant, potato and *S. nigrum*) but also with additional species of Iris, *Ipompoea indivisia,* and *Silene* dioica (Table 3). These results emphasised the conservation of metabolic decoration of anthocyanins, in the form of 3-*O*-(4'''-phenylacyl-rutinoside)-5-*O*-glucoside derivatives in Solanaceous species. In spite of this metabolite conservation of anthocyanin-3-*O*-(4'''-phenylacyl-rutinoside)-5-*O*-glucoside derivatives, flavonol derivatives with 3-*O*-(4'''pheynlacyl)-rutinoside decorations were limited to kaempferol-3-*O*-(4'''-(3'''-*O*-rhamnosyl)-*trans*-*p*-coumaroyl-rutinoside in *Dicranopteris linearis*. Although flavonol-3-*O*-rutinosides are one of the most abundant flavonols in the plant kingdom, their acylated derivatives appear to be uncommon in plants.

Phylogenetic analysis of candidate genes involved in anthocyanin production in tomato

On the basis of the phenylacyl decoration of the major anthocyanins abundant in the purple tomato lines and their conservation within the Solanaceae it seemed appropriate to perform a phylogenetic analysis of candidate genes responsible for this specific decoration. Given the characterized chemical structures of the decorated anthocyanins and flavonols and RNAseq data, four of the genes (three UDP-glycosyltransferase genes: Solyc09g059170, Solyc10g083440 and Solyc12g098590, and one BEAT/AHCT/HCBT/DAT (BAHD) gene: Solyc12g088170) which were observed to be upregulated at the transcript level in purple tomatoes represent very strong candidates for being involved in anthocyanin decoration in tomato. Flavonoid-glycosyltransferase genes most commonly belong to the UGT1 gene family. The phylogenetic analysis shown in Figure 4A demonstration that the candidate anthocyanidin-glycosyltransferases from tomato fall into distinct clades within the UGT1 family, with the GTs from tomato being most similar in each case to one from

petunia. The candidate genes encoded proteins belonging to three different classes, namely a gene encoding a flavonoid-3-*O*-glycosyltransferase (SlFd3GlcT; Solyc10g083440), one encoding an anthocyanin-5-*O*-glycosyltransferase (SlA5GlcT; Solyc12g098590) and one encoding a flavonoid-3-*O*-glycoside-*O*-glycosyltransferase (SlA3Glc6''RhaT; Solyc09g059170). This sequence similarity to other genes from both within the Solanaceous family and from other plant families suggested that the anthocyanin sugar-transferases evolved prior to the divergence of Solanaceous species. In contrast, the gene encoding a putative flavonoid-acyltransferase, Solyc12g088170 (which we named SlFdAT1), did not cluster with any previously characterized anthocyanin-acyltransferase (Figure 4B) meaning that, it could be classified only as a member of the BAHD family (reviewed in D'Auria, 2006) on the basis of phylogeny. All these genes encoding candidate decorating enzymes showed similar expression patterns in specific tissues of tomato in eFP (Supplemental Figure S7).

As a further step towards understanding the anthocyanin metabolic pathway we searched the PLAZA database (http://bioinformatics.psb.ugent.be/plaza) for orthologs of the tomato genes known to, or suggested from our study to encode each step of the anthocyanin biosynthetic pathway. The results of this search are presented in Figure 4C wherein for each enzyme the number of genes found and the number of species in which they were found. For the majority of enzymatic steps genes were found in excess of 20 species and, for some, enzymatic reactions are obviously catered for by a large number of isoforms. Four exceptions containing F3'5'H, A3GlcT, A3Glc6''RhaT and SlFdAT1 are noteworthy (Figure 4C). Since SlFdAT1 shows very poor conservation and is limited to three species, we next performed a search for orthologues of SlFdAT using gene sequences submitted to GeneBank (Figure 4D). Targeted phylogenetic analysis resulted in clear separation between a BAHD subclade which included SlFdAT1 and other ATs from the Solanaceae, such as potato and tobacco, but did not include ATs from other plant families.

SlFdAT1 **encodes a flavonoid-3-*O*-rutinoside-4'''-*O*-phenylacyl transferase**

To characterize the biochemical function of SlFdAT1, assays of the recombinant enzyme were performed (Figure 5A-D). The cDNA sequence of Solyc12g088170

annotated by genome annotation in SOL was amplified and recombined into expression vector pJAM1784 (Luo et al., 2007) to produce pJAM1786 which expresses SlFdAT fused N-terminally to the S-TAG of RNase S. The recombinant protein was expressed in *E. coli*, and total protein was measured in crude extracts by S-TAG Rapid Assay Kit (Novagen). The activity of SlFdAT1 with six acyl-donors (*p*-coumaroyl-CoA, feruloyl-CoA, caffeoyl-CoA, sinapoyl-CoA, cinnamoyl-CoA and malonyl-CoA) and four acyl-acceptors (cyanidin-3-*O*-rutinoside, quercetin-3-*O*-rutinoside and kaempferol-3-*O*-rutinoside and delphinidin-3-*O*-rutinoside) were tested. We were unable to prepare large enough amounts of substrates for determination of complete kinetic parameters, but we were able to characterize substrate specificities based on relative activities. Reaction products were confirmed by LC/PDA-MS. The recombinant enzyme showed -4'''-*O*-phenylacyltransferase activity with both anthocyanin (Figures 5A and B) and flavonol acyl-acceptors (Figures 5C and D) and all phenylacyl-donors, but had no activity with malonyl-CoA as the acyl-donor (Table 4).

Additionally, 35S-driven *SlFdAT1* was overexpressed in tobacco to examine SlFdAT1 function in vivo (Figures 5E-G). In tobacco, anthocyanins accumulate in the flower limb exclusively. These are largely cyanidin-3-*O*-rutinoside, with small amounts of pelargonidin-3-*O*-rutinoside. There are no reports of acylated anthocyanins from tobacco. Metabolite profiling of flower extracts of a *SlFdAT1*-overexpressing line showed production of significant levels of phenylacylated-cyanidin-3-*O*-rutinosides and trace levels of phenylacylated-quercetin-3-*O*-rutinoside (Figures 5F and G and Supplemental Figure S8). Given the results of the assays of recombinant Sl*Fd*AT1 protein in vitro, these data confirmed that Sl*Fd*AT1 acts as a flavonoid-3-*O*-rutinoside-4'''-*O*-phenylacyltransferase in vivo, and its activity can account for the acylation of both anthocyanins and flavonols observed in tomato. Taken together, the phylogenetic analyses suggest that SlFdAT1 evolved within the Solanaceae, independently of other anthocyanin-phenylacyltransferases, implying that the convergent evolution of SlFdAT1 to adopt anthocyanin phenylacyl-transferase activity, means that its broad substrate specificity for hydroxycinnamoyl-CoA acyl-donors and for both flavonol and anthocyanin acyl-acceptors is relatively unusual in the plant kingdom.

DISCUSSION

The overexpression of transcription factors has proven a highly effective strategy to ascertain their function (Tamagnone et al., 1998; Tohge et al., 2005; Wu et al., 2012; Schmidt et al., 2013), and to facilitate identification of the metabolic pathways which they regulate (Sharma and Dixon, 2005; Tohge et al., 2005; Peel et al., 2009; Zhao et al., 2010; Liu et al., 2014). Here we carried out an in depth analysis of a *Del/Ros1* purple tomatoes that accumulate anthocyanins at concentrations comparable to those found in blackberries and blueberries, enhancing the antioxidant potential of the fruit by threefold and extending the lifespan of cancer-susceptible mice in a dietary context (Butelli et al., 2008). The presence of such high levels of anthocyanins in tomatoes was subsequently demonstrated to double the shelf life of the fruit by delaying over-ripening and reducing susceptibility to *Botrytis cinerea* (Zhang et al., 2013). Our current study utilized comprehensive metabolite profiling alongside RNA-seq to gain a fuller picture of the compositional changes invoked by fruit-specific, ectopic expression of these transcription factors. In addition, using the E8 promoter to switch on anthocyanin biosynthesis is equivalent to using an inducible promoter (albeit a slow induction), and analysis of the effects of the transcription factors is more likely to reveal direct targets and less likely to be compromised by indirect effects resulting from use of constitutive promoters. We showed that over-expression of *Del/Ros1* resulted in multiple changes in addition to the changes in anthocyanins reported in previous studies (Butelli et al., 2008; Zhang et al., 2013; Zhang et al., 2014). These changes included depletion of Phe, rutin and naringenin-chalcone levels but enhanced levels of total carotenoids, CGAs and flavonol-derivatives. The accumulated anthocyanins were quite unusual, with phenylacylation of the rhamnose sugar of the rutinoside.

While the metabolic changes in the *Del/Ros1* transgenic line were relatively focussed on anthocyanins, several of the other changes are worthy of discussion, particularly when these data are evaluated in conjunction with those from RNA-seq analysis. One prominent change was the reduced levels of Phe, which was observed in peel, pericarp and whole fruit samples, as were decreased levels of alanine, threitol and galactose. Phe is the direct precursor of the phenylpropanoid pathway and its decreased levels are likely the direct result of increased demand for anthocyanin biosynthesis. Interestingly, analysis of the RNA-seq data revealed that

this decrease occurred in spite of an elevated expression of the genes of Phe biosynthesis, suggesting that the increased demand for Phe in the purple tomatoes outstrips the predicted enhanced production of the amino acid. On the other hand, quinate and several isoforms of CGAs were increased in the *Del*/*Ros1* line. Since RNAseq analysis showed up-regulation of genes involved in not only flavonoid but also phenylpropanoid, phenylalanine and shikimate biosynthesis, over-accumulation of quinate and CGAs are likely the result of enhanced demand for the production of flavonoids in the *Del/Ros1* line.

Structural elucidation of tomato anthocyanins has previously proven very difficult, since they are not usually available in large amounts (Jones et al., 2003) and prior to this, accumulation as a result of genetic manipulation had rarely been detected in tomato flesh (Bovy et al., 2002). A low abundance of anthocyanins in general tissues is observed in tomatoes and in most non-domesticated tomato accessions such as *S. pennellii*. Using the high-yielding purple tomatoes, 6 major anthocyanins were identified which include the same anthocyanins found in eggplant and petunia. Co-elution profiles of anthocyanins from *Del/Ros1* fruits and M82/*pennellii* leaves revealed that these anthocyanins are the same as endogenous anthocyanins in tomato species.

Genome scale profiling of anthocyanin biosynthesis has already been reported for Arabidopsis using the mutant *pap1*-D which overexpresses the *AtMYB75/PAP1* gene and induces over expression of bHLH42 (AtTT8) which is up-regulated by AtMYB75 (Tohge et al., 2005). We performed cross species comparisons of up-regulated genes in high anthocyanin mutants and transgenic plants (Figure 6). This approach suggested conservation of the transcriptional networks which enable up-regulation of genes involved in the first steps of the anthocyanin biosynthetic pathway (EBGs) which are common to most Angiosperms, and some late biosynthetic genes (LBGs) namely 3-*O*-glucosyl and 5-*O*-glucosyltransferase genes. This is in contrast to reports of the levels of target gene expression in mutants affected in the activity of the regulatory genes (Quattrocchio et al., 1993; Winkel-Shirley, 2001; Schwinn et al., 2006; Purdy et al., 2013). In such mutants, however, expression of EBGs may be complemented by the activity of other MYB regulators, such as members of the MYB family, subgroup 7. This means that

EBGs may be mis-scored as not regulated by the MBW complex, of which Del and Ros1 are part (Zhang et al., 2014).

Since anthocyanin acyl-acceptors are not exactly the same in Arabidopsis (anthocyanin-3-*O*-glucoside) and Solanaceae species (anthocyanin-3-*O*-rutinoside), differences could be observed in LBGs of glycosyltransferases and acyltransferases for the two species. The dual functionality of SlFdAT1 in catalyzing the transfer of caffeoyl, *p*-coumaroyl or feruloyl moieties, to either a flavonol or an anthocyanin in *Del/Ros1* tomato appears to be unique amongst characterized BAHD acyltransferases (Table 4 and Figure 5). For example, in Arabidopsis, the enzyme that acylates anthocyanin-3-*O*-glucoside is specific for anthocyanins as acyl-acceptors (Luo et al., 2007). Further cross-species comparisons suggested that the anthocyanins characterized in this study are conserved among Solanaceae species. Although phenylacylated-anthocyanidin-rutinosides are found in a few other distantly related plant species, it is likely that in these species decorating enzymes (particularly acyltransferases) have evolved by convergence and consequently may have rather different properties to SlFdAT1 and the other anthocyanin acyltransferases conserved in Solanaceous species (Nunes-Nesi et al., 2014). Unusually, several phenylacylated-flavonol-derivatives accumulated in *Del/Ros1* purple tomatoes. Interestingly, although flavonol-3-*O*-rutinosides are one of the common flavonol-glycosides (Table 3), flavonol-3-*O*-(4'''-phenylacylated)-rutinosides which are products of SlFdAT1are very uncommon in the plant kingdom. SlFdAT1 can accept several acyl-donors (caffeoyl, *p*-coumaroyl, cinnamoyl, sinapoyl or feruloyl-CoAs) and use anthocyanin-3-*O*-rutinosides and also flavonol-3-*O*-rutinosides as acyl-acceptors.

This analysis showed that ectopic expression of two components of the MBW complex (Ros1/[Myb] and Del/[bHLH]) resulted in up-regulated expression of genes involved in phenylpropanoid metabolism, flavonoid/anthocyanin biosynthesis and anthocyanin decoration. It is likely, based on the inducible nature of expression of Del/Ros1 by the E8 promoter, that all genes with very significantly enhanced expression are direct targets of the transcription factors. Comparison to metabolomic and transcriptomic data from the *pap1*-D mutant of Arabidopsis indicates that *Del/Ros1* in tomato switch on the same genes of general phenylpropanoid metabolism and flavonoid biosynthesis as PAP1 does together with bHLH

transcription factors (GL3, EGL3, AtTT8) in Arabidopsis (Tohge et al., 2005). Differences in induced expression of genes encoding anthocyanin decorating enzymes were observed between tomato and Arabidopsis, but this reflects the different decoration of anthocyanins in these two species. Interestingly, *Del/Ros1* induced expression of genes early in the flavonoid biosynthetic pathway as well as later steps, implying that these transcription factors turn on all the genes required for conversion of *p*-coumaroyl-CoA to anthocyanins. Amongst the genes encoding decorating enzymes those encoding glycosyltransferases are highly conserved with those from other Angiosperm species implying that A3GlcT, A5GlcT and A3Glc6''RhaT evolved relatively early before the divergence of the family Solanaceae. In Solanaceous species such as tobacco which lacks 5-*O*-glycosylation of its anthocyanins, the gene encoding the A5GlcT likely has been lost. In contrast, the gene encoding SlFdAT1 appears to have arisen after the divergence of the Solanaceae, and must have recruited the MBW to control its expression, once its specificity had been determined. In addition, the specificity of this enzyme, particularly for acyl acceptors, may be useful in terms of engineering new colours and other functionalities of flavonoids, in the future.

EXPERIMENTAL PROCEDURES

Plant growth

The generation and molecular characterization of the transgenic plant material has been described in detail previously (Butelli et al., 2008; Zhang et al., 2013). Tomato plants, MicroTom, Del/Ros1 line N in MicroTom, M82 and *S. pennellii*, were handled as described previously in the literature (Kochevenko et al., 2012). Whole fruits and two tissues (flesh and peel) were harvested at two stages (Breaker + 1week and + 4weeks)(Zhang et al., 2013) and used for the analysis. Four and two biological replicates harvested from individual plants were used for metabolite profiling and RNA-seq expression analysis, respectively.

Profiling of primary and secondary metabolites

Non-targeted metabolite profiling was carried out by GC-MS following exactly the protocol of Lisec et al., (Lisec et al., 2006) and by LC-MS by extracting metabolites following the protocol described in Tohge et al. (2010) and analysing these extracts

exactly as described in Rohrmann et al., (2011) and Schwahn et al., (2014). Information concerning metabolite identification is providing following reporting suggestions for large-scale metabolite datasets (Supplemental Table S1)(Fernie et al., 2011).

RNA sequencing

EXPRESS Tag sequencing was performed as described by Rallapalli et al. (2014). Briefly, total RNA (3 µg) was used to generate first strand cDNA using an oligo (dT) primer comprising P7 sequence of Ilumina flow cell. Double strand cDNA was synthesized as described previously (Okayama and Berg, 1982). Purified cDNA was subjected to Covaris shearing (parameters: Intensity 5, Duty cycle 20%, Cycles/Burst 200, Duration 90 seconds). End repairing and A-tailing of sheared cDNA was carried out as described by Illumina. Y-shaped adapters were ligated to A-tailed DNA and subjected to size selection on agarose gels. The gel-extracted library was PCR enriched and quantified using qPCR with previously sequenced similar size range Illumina libraries. The libraries were sequenced on Illumina Genome Analyzer IIx. The Illumina sequence library was quality filtered using FASTX Toolkit 0.0.13 with parameters -q20 and -p50 (http://hannonlab.cshl.edu/fastx_toolkit/index.html). The sub-library was artefact-filtered using the FASTX-toolkit. Quality filtered libraries were aligned to the *S. lycopersicum* cDNA sequences (ITAG2.3 ftp://ftp.solgenomics.net/tomato_genome/annotation/ITAG2.3_release/ITAG2.3_cdna .fasta) using Bowtie version 0.12.8 (Langmead et al., 2009). Unaligned reads from the previous step were used to align to Tomato genome: (ftp://ftp.solgenomics.net/tomato_genome/annotation/ITAG2.3_release/ITAG2.3_gen omic.fasta) using Bowtie version 0.12.8. The RNA-seq data has been deposited in NCBI's Gene Expression Omnibus (GEO) with series accession number GSE61014.

Isolation and purification of the anthocyanins

For the isolation of the major anthocyanins in *Del/Ros1* tomato fruit, extracts were prepared from 5 g powdered tomato fruit with 50 ml of 50% aqueous MeOH and 25 ml of 100% MeOH. The clarified extract by paper filtration was then concentrated in a rotary evaporator to a final volume of 20 ml and purified by preparative HPLC. HPLC condition was described in Supplemental file 1.

Identification of acyl-moieties sugar-moieties by hydrolysis

For the identification of acyl-moieties, 20 µl of pure anthocyanin fractions were subjected to alkaline hydrolysis with 250 µl 10% KOH for 30 min at room temperature. The hydrolysate was acidified to pH 1.0 with 250 µl 2 N HCl and the saponification products were extracted with three volumes ethyl acetate. For identification of sugar-moieties, 20 µl of pure anthocyanin fractions were subjected to acid hydrolysis with 120 µl 2 M HCl for 30 min at 95°C in a sealed vial. After cooling of the hydrolysate in an ice bath were extracted with 1 ml 1-pentanol. Obtained fractions were dried and resuspended for subjecting to LC/MS. LC/MS method for profiling of hydrolysate was described in Supplemental file 1.

Nuclear magnetic resonance to identify anthocyanins

^{1}H and ^{13}C NMR spectra were acquired with a Bruker DMX 500 NMR spectrometer (Rheinstetten, Germany) operating at 500.13 MHz for ^{1}H and 125.77 MHz for ^{13}C. CD_3OD/CF_3COOD (10:1) solutions of ~1 mg sample were run at 303 K under standard conditions and pulse sequences (^{1}H-NMR; ^{13}C-NMR; ^{13}C-DEPT; gs-2Q-^{1}H, ^{1}H-COSY; gs-^{1}H, ^{13}C-HSQC; gs-^{1}H, ^{13}C-HMBC; gs-^{1}H, ^{1}H-NOESY; gs-^{1}H, ^{13}C-TOCSY-HSQC) as previously described (Kaffarnik et al., 2005). Reference chemical shifts were 3.30 ppm and 49.00 ppm for residual C**H**D_2OD and **C**D_3OD signals, respectively. ^{1}H and ^{13}C NMR spectral data is shown in Supplemental file S2.

In vitro assay of recombinant SlFdAT1 protein

Enzymatic assay of SlFdAT1 (Solyc12g088170) using recombinant protein was performed by a modification of the method described in Luo et al (2007). Full length *SlFdAT1* (Solyc12g088170) cDNA was amplified from *Del/Ros1* tomato fruit cDNA using the following primers:

SlFdAT1-attB1:
GGGGACAAGTTTGTACAAAAAAGCAGGCTGGATGAGCCAAATTACAACACAAAA (5'-3')

SlFdAT1-attB2:
GGGGACCACTTTGTACAAGAAAGCTGGGTCCTACTACTTTGGCACATAACTA (5'-3').

PCR products were recombined with pDONR207 vectors to make pENTR207-*SlFdAT1*. After sequence verification, the entry clone was introduced into the N-terminal S-TAG fusion vector pJAM1784 to make pJAM1784-*SlFdAT1*. Recombinant protein expression and purification was done as described previously (Luo et al., 2007). The S-TAG protein concentration was determined using an S-TAG Rapid Assay Kit (Novagen) following the manufacturer's instructions. Enzyme assays were carried as described previously (Luo et al., 2007).

Acyl-donors for in vitro assays, *p*-coumaroyl-CoA, feruloyl-CoA, caffeoyl-CoA, sinapoyl-CoA and cinnamoyl-CoA, were purchased from TransMIT GmbH, Project Division for Plant Metabolites and Chemicals (PMC). Quercetin-3-*O*-rutinoside (rutin) and kaempferol-3-*O*-rutinoside were purchased from Sigma. Delphinidin-3-*O*-rutinoside was purchased from EXTRASYNTHESE (Lyon, France). LC/MS method used for the detection of enzymatic products is described in Supplemental file 3.

Ectopic Expression of SlFdAT1 in Tobacco

The *SlFdAT1* (Solyc12g088170) full length cDNA was amplified with primers SlFdAT1-attB1 and SlFdAT1-attB2 and cloned in a pBin19-GW binary vector (Butelli et al., 2012) through Gateway cloning. The resulting plasmid was transferred to *Agrobacterium tumefaciens* strain LBA4404 and used to transform tobacco (*Nicotiana tabacum cv Samsun*) as described previously (Luo et al., 2008). Leaves of transgenic line were harvested and applied to LC/MS analysis.

ACKNOWLEDGMENTS

T.T and A.F. gratefully acknowledge partial support by the Max Planck Society and an Alexander von Humboldt grant (to T.T.) as well as the European Commission's Directorate-General for Research within the 7th Framework Program (FP7/2007-2013) under grant agreements 270089 (MULTIBIOPRO). The authors wish to acknowledge EU 6th Framework project 007130 "FLORA". We would like to thank to Ms. Yi Tse Liu for the excellent technical support.

Supporting legends

Supplemental Figure S1. LC/MS total ion chromatogram (TIC, negative ion detection) of WT and *Del/Ros1* tomatoes in flesh and peel.

Supplemental Figure S2. Changes in primary and secondary metabolites in *Del/Ros1* tomatoes compared to WT.

Supplemental Figure S3. Global overview of transcriptome data of metabolism-related genes developed using Mapman.

Supplemental Figure S4. Semipreparative HPLC profile at 535 nm of the main anthocyanins from the *Del/Ros1* tomato transgenic tomato.

Supplemental Figure S5. Identification of petunidin-3-*O*-[4-(*trans*-*p*-coumaroyl)-rutinoside]-5-*O*-*β*-D-glucoside (anthocyanin TA1).

Supplemental Figure S6. Expression data analysis of anthocyanin biosynthesis related genes in *S. lycopersicum* and *S. pennellii* tissues.

Supplemental Figure S7. eFP gene expression analysis.

Supplemental Figure S8. MSMS analysis of major anthocyanins found in *35S:SlFdAT1* transgenic tobacco flower.

Supplemental Table S1. Summary of values for secondary metabolites (glycoalkaloids, flavonoids and phenylpropanoids) for WT and *Del/Ros1* tomato.

Supplemental Table S2. RNAseq data of WT and Del/Ros1 transgenic tomato line.

Supplemental File S1. HPLC methods for Preparative HPLC method for purification of the anthocyanins from *Del/Ros1* tomato transgenic line, and for profiling of hydrolysate.

Supplemental File S2. ^{1}H and ^{13}C NMR spectral data of compounds III and V [in CD_3OD/CF_3COOD (10:1) at 500.13 MHz].

Supplemental File S3. LC/MS method used for the detection of enzymatic products.

REFFERENCE

A. B, R. dV, M. K, E. S, A. PM, S. M, G. C, S. R, M. V, S. H, C. S-B, A. vT (2002) High-flavonol tomatoes resulting from the heterologous expression of the maize transcription factor genes LC and C1. Plant Cell **14:** 2509-2526

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

Alseekh S, Tohge T, Wendenberg R, Omranian N, Kleessen S, Giavalisco P, Pleban T, Mueller-Roeber B, Zamir D, Nikoloski Z, Fernie A, R. (2015) Identification and mode of inheritance of quantitative trait loci for secondary metabolite abundance in tomato. Plant Cell**:** in-presss

Aviram M, Fuhrman B (2002) Wine flavonoids protect against LDL oxidation and atherosclerosis. *In* DK Das, F Ursini, eds, Alcohol and Wine in Health and Disease, Vol 957, pp 146-161

Bolger A, Scossa F, Bolger ME, Lanz C, Maumus F, Tohge T, Quesneville H, Alseekh S, Sorensen I, Lichtenstein G, Fich EA, Conte M, Keller H, Schneeberger K, Schwacke R, Ofner I, Vrebalov J, Xu Y, Osorio S, Aflitos SA, Schijlen E, Jimenez-Gomez JM, Ryngajllo M, Kimura S, Kumar R, Koenig D, Headland LR, Maloof JN, Sinha N, van Ham RCHJ, Lankhorst RK, Mao L, Vogel A, Arsova B, Panstruga R, Fei Z, Rose JKC, Zamir D, Carrari FF, Giovannoni JJ, Weigel D, Usadel B, Fernie AR (2014) The genome of the stress-tolerant wild tomato species Solanum pennellii. Nature Genetics **46:** 1034-+

Borevitz JO, Xia YJ, Blount J, Dixon RA, Lamb C (2000) Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. Plant Cell **12:** 2383-2393

Bovy A, de Vos R, Kemper M, Schijlen E, Pertejo MA, Muir S, Collins G, Robinson S, Verhoeyen M, Hughes S, Santos-Buelga C, van Tunen A (2002) High-flavonol tomatoes resulting from the heterologous expression of the maize transcription factor genes LC and C1. Plant Cell **14:** 2509-2526

Brechenmacher L, Lei ZT, Libault M, Findley S, Sugawara M, Sadowsky MJ, Sumner LW, Stacey G (2010) Soybean Metabolites Regulated in Root Hairs in Response to the Symbiotic Bacterium Bradyrhizobium japonicum. Plant Physiology **153:** 1808-1822

Butelli E, Licciardello C, Zhang Y, Liu J, Mackay S, Bailey P, Reforgiato-Recupero G, Martin C (2012) Retrotransposons Control Fruit-Specific, Cold-Dependent Accumulation of Anthocyanins in Blood Oranges. Plant Cell **24:** 1242-1255

Butelli E, Titta L, Giorgio M, Mock HP, Matros A, Peterek S, Schijlen EGWM, Hall RD, Bovy AG, Luo J, Martin C (2008) Enrichment of tomato fruit with health-

promoting anthocyanins by expression of select transcription factors. Nature Biotechnology **26:** 1301-1308

Carrari F, Baxter C, Usadel B, Urbanczyk-Wochniak E, Zanor MI, Nunes-Nesi A, Nikiforova V, Centero D, Ratzka A, Pauly M, Sweetlove LJ, Fernie AR (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

Commenges D, Scotet V, Renaud S, Jacqmin-Gadda H, Barberger-Gateau P, Dartigues JF (2000) Intake of flavonoids and risk of dementia. European Journal of Epidemiology **16:** 357-363

D'Auria JC (2006) Acyltransferases in plants: a good time to be BAHD. Current Opinion in Plant Biology **9:** 331-340

Dal Cin V, Tieman DM, Tohge T, McQuinn R, de Vos RC, Osorio S, Schmelz EA, Taylor MG, Smits-Kroon MT, Schuurink RC, Haring MA, Giovannoni J, Fernie AR, Klee HJ (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

Enfissi EM, Barneche F, Ahmed I, Lichtle C, Gerrish C, McQuinn RP, Giovannoni JJ, Lopez-Juez E, Bowler C, Bramley PM, Fraser PD (2010) Integrative transcript and metabolite analysis of nutritionally enhanced DE-ETIOLATED1 downregulated tomato fruit. Plant Cell **22:** 1190-1215

Fernie AR, Aharoni A, Willmitzer L, Stitt M, Tohge T, Kopka J, Carroll AJ, Saito K, Fraser PD, DeLuca V (2011) Recommendations for Reporting Metabolite Data. Plant Cell **23:** 2477-2482

Fitzpatrick TB, Basset GJC, Borel P, Carrari F, DellaPenna D, Fraser PD, Hellmann H, Osorio S, Rothan C, Valpuesta V, Caris-Veyrat C, Fernie AR (2012) Vitamin Deficiencies in Humans: Can Plant Science Help? Plant Cell **24:** 395-414

Fraser CM, Chapple C (2011) The phenylpropanoid pathway in Arabidopsis. The Arabidopsis book / American Society of Plant Biologists **9:** e0152-e0152

Grotewold E, Sainz MB, Tagliani L, Hernandez JM, Bowen B, Chandler VL (2000) Identification of the residues in the Myb domain of maize C1 that specify the interaction with the bHLH cofactor R. Proceedings of the National Academy of Sciences of the United States of America **97:** 13579-13584

Gutteridge JMC (1994) BIOLOGICAL ORIGIN OF FREE-RADICALS, AND MECHANISMS OF ANTIOXIDANT PROTECTION. Chemico-Biological Interactions **91:** 133-140

Halliwell B (2012) Free radicals and antioxidants: updating a personal view. Nutrition Reviews **70:** 257-265

Handa AK, Mattoo AK (2010) Differential and functional interactions emphasize the multiple roles of polyamines in plants. Plant Physiology and Biochemistry **48:** 540-546

Harborne JB (1964) PLANT POLYPHENOLS .11. THE STRUCTURE OF ACYLATED ANTHOCYANINS. Phytochemistry **3:** 151-160

Hirai MY, Sugiyama K, Sawada Y, Tohge T, Obayashi T, Suzuki A, Araki R, Sakurai N, Suzuki H, Aoki K, Goda H, Nishizawa OI, Shibata D, Saito K (2007) Omics-based identification of Arabidopsis Myb transcription factors regulating aliphatic glucosinolate biosynthesis. Proceedings of the National Academy of Sciences of the United States of America **104:** 6478-6483

Hollman PCH, Hertog MGL, Katan MB (1996) Analysis and health effects of flavonoids. Food Chemistry **57:** 43-46

Huang J, Gu M, Lai Z, Fan B, Shi K, Zhou YH, Yu JQ, Chen Z (2010) Functional analysis of the Arabidopsis PAL gene family in plant growth, development, and response to environmental stress. Plant Physiology **153:** 1526

Jones CM, Mes P, Myers JR (2003) Characterization and inheritance of the Anthocyanin fruit (Aft) tomato. Journal of Heredity **94:** 449-456

Kaffarnik F, Heller W, Hertkorn N, Sandermann H (2005) Flavonol 3-O-glycoside hydroxycinnamoyltransferases from Scots pine (Pinus sylvestris L.). Febs Journal **272:** 1415-1424

Kochevenko A, Klee HJ, Fernie AR, Araujo WL (2012) Molecular identification of a further branched-chain aminotransferase 7 (BCAT7) in tomato plants. Journal of Plant Physiology **169:** 437-443

Koenig D, Jimenez-Gomez JM, Kimura S, Fulop D, Chitwood DH, Headland LR, Kumar R, Covington MF, Devisetty UK, Tat AV, Tohge T, Bolger A, Schneeberger K, Ossowski S, Lanz C, Xiong G, Taylor-Teeples M, Brady SM, Pauly M, Weigel D, Usadel B, Fernie AR, Peng J, Sinha NR, Maloof JN (2013) Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. Proceedings of the National Academy of Sciences of the United States of America **110:** E2655-E2662

Kong Q, Pattanaik S, Feller A, Werkman JR, Chai C, Wang Y, Grotewold E, Yuan L (2012) Regulatory switch enforced by basic helix-loop-helix and ACT-domain mediated dimerizations of the maize transcription factor R. Proceedings of the National Academy of Sciences of the United States of America **109:** E2091-E2097

Kusano M, Tohge T, Fukushima A, Kobayashi M, Hayashi N, Otsuki H, Kondou Y, Goto H, Kawashima M, Matsuda F, Niida R, Matsui M, Saito K, Fernie AR (2011) Metabolomics reveals comprehensive reprogramming involving two independent metabolic responses of Arabidopsis to UV-B light. Plant Journal **67:** 354-369

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol **10:** R25

Le Gall G, DuPont MS, Mellon FA, Davis AL, Collins GJ, Verhoeyen ME, Colquhoun IJ (2003) Characterization and content of flavonoid glycosides in genetically modified tomato (Lycopersicon esculentum) fruits. Journal of Agricultural and Food Chemistry **51:** 2438-2446

Le Marchand L (2002) Cancer preventive effects of flavonoids - a review. Biomedicine & Pharmacotherapy **56:** 296-301

Lepelley M, McCarthy JG, Petiard V, Cheminade G, Tanksley SD, Lin C (2012) Polynucleotides encoding lignin biosynthetic pathway enzymes in coffee. *In*. Nestec S A; Cornell University, USA

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

Liu Z, Shi M-Z, Xie D-Y (2014) Regulation of anthocyanin biosynthesis in Arabidopsis thaliana red pap1-D cells metabolically programmed by auxins. Planta **239:** 765-781

Luo J, Butelli E, Hill L, Parr A, Niggeweg R, Bailey P, Weisshaar B, 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 Journal **56:** 316-326

Luo J, Fuell C, Parr A, Hill L, Bailey P, Elliott K, Fairhurst SA, Martin C, Michael AJ (2009) A Novel Polyamine Acyltransferase Responsible for the

Accumulation of Spermidine Conjugates in Arabidopsis Seed. Plant Cell **21:** 318-333

Luo J, Nishiyama Y, Fuell C, Taguchi G, Elliott K, Hill L, Tanaka Y, Kitayama M, Yamazaki M, Bailey P, Parr A, Michael AJ, Saito K, Martin C (2007) Convergent evolution in the BAHD family of acyl transferases: identification and characterization of anthocyanin acyl transferases from Arabidopsis thaliana. Plant Journal **50:** 678-695

Martin C (2013) The interface between plant metabolic engineering and human health. Current Opinion in Biotechnology **24:** 344-353

Martin C, Zhang Y, Tonelli C, Petroni K (2013) Plants, Diet, and Health. Annual Review of Plant Biology, Vol 64 **64:** 19-46

Mathews H, Clendennen SK, Caldwell CG, Liu XL, Connors K, Matheis N, Schuster DK, Menasco DJ, Wagoner W, Lightner J, Wagner DR (2003) Activation tagging in tomato identifies a transcriptional regulator of anthocyanin biosynthesis, modification, and transport. Plant Cell **15:** 1689-1703

Mojzisova G, Kuchta M (2001) Dietary flavonoids and risk of coronary heart disease. Physiological Research **50:** 529-535

Mounet F, Moing A, Garcia V, Petit J, Maucourt M, Deborde C, Bernillon S, Le Gall G, Colquhoun I, Defernez M, Giraudel JL, Rolin D, Rothan C, Lemaire-Chamley M (2009) Gene and Metabolite Regulatory Network Analysis of Early Developing Fruit Tissues Highlights New Candidate Genes for the Control of Tomato Fruit Composition and Development. Plant Physiology **149:** 1505-1528

Muir SR, Collins GJ, Robinson S, Hughes S, Bovy A, De Vos CHR, van Tunen AJ, Verhoeyen ME (2001) Overexpression of petunia chalcone isomerase in tomato results in fruit containing increased levels of flavonols. Nature Biotechnology **19:** 470-474

Nerdal W, Andersen OM (1992) INTERMOLECULAR AROMATIC ACID ASSOCIATION OF AN ANTHOCYANIN (PETANIN) EVIDENCED BY 2-DIMENSIONAL NUCLEAR OVERHAUSER ENHANCEMENT NUCLEAR-MAGNETIC-RESONANCE EXPERIMENTS AND DISTANCE GEOMETRY CALCULATIONS. Phytochemical Analysis **3:** 182-189

Nunes-Nesi A, Florian A, Howden A, Jahnke K, Timm S, Bauwe H, Sweetlove L, Fernie AR (2014) Is There a Metabolic Requirement for Photorespiratory Enzyme Activities in Heterotrophic Tissues? Molecular Plant **7:** 248-251

Okayama H, Berg P (1982) High-efficiency cloning of full-length cDNA. Mol Cell Biol **2:** 161-170

Peel GJ, Pang Y, Modolo LV, Dixon RA (2009) The LAP1 MYB transcription factor orchestrates anthocyanidin biosynthesis and glycosylation in Medicago. Plant Journal **59:** 136-149

Purdy SJ, Bussell JD, Nunn CP, Smith SM (2013) Leaves of the Arabidopsis maltose exporter1 Mutant Exhibit a Metabolic Profile with Features of Cold Acclimation in the Warm. Plos One **8:** e79412

Quattrocchio F, Wing JF, Leppen HTC, Mol JNM, Koes RE (1993) REGULATORY GENES-CONTROLLING ANTHOCYANIN PIGMENTATION ARE FUNCTIONALLY CONSERVED AMONG PLANT-SPECIES AND HAVE DISTINCT SETS OF TARGET GENES. Plant Cell **5:** 1497-1512

Rallapalli G, Kemen EM, Robert-Seilaniantz A, Segonzac C, Etherington GJ, Sohn KH, MacLean D, Jones JDG (2014) EXPRSS: an Illumina based high-

throughput expression-profiling method to reveal transcriptional dynamics. Bmc Genomics **15**

Ramsay NA, Glover BJ (2005) MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. Trends in Plant Science **10:** 63-70

Rohrmann J, Tohge T, Alba R, Osorio S, Caldana C, McQuinn R, Arvidsson S, van der Merwe MJ, Riaño-Pachón DM, Mueller-Roeber B, Fei Z, Nesi A, Giovannoni J, Fernie A (2011) Combined transcription factor profiling, microarray analysis and metabolite profiling reveals the transcriptional control of metabolic shifts occurring during tomato fruit development. Plant Journal **68:** 999-1013

Schmidt R, Mieulet D, Hubberten H-M, Obata T, Hoefgen R, Fernie AR, Fisahn J, Segundo BS, Guiderdoni E, Schippers JHM, Mueller-Roeber B (2013) SALT-RESPONSIVE ERF1 Regulates Reactive Oxygen Species-Dependent Signaling during the Initial Response to Salt Stress in Rice. Plant Cell **25:** 2115-2131

Schwahn K, de Souza LP, Fernie AR, Tohge T (2014) Metabolomics-assisted refinement of the pathways of steroidal glycoalkaloid biosynthesis in the tomato clade. Journal of Integrative Plant Biology **56:** 864-875

Schwinn K, Venail J, Shang YJ, Mackay S, Alm V, Butelli E, Oyama R, Bailey P, Davies K, Martin C (2006) A small family of MYB-regulatory genes controls floral pigmentation intensity and patterning in the genus Antirrhinum. Plant Cell **18:** 831-851

Sharma SB, Dixon RA (2005) Metabolic engineering of proanthocyanidins by ectopic expression of transcription factors in Arabidopsis thaliana. Plant Journal **44:** 62-75

Sonderby IE, Hansen BG, Bjarnholt N, Ticconi C, Halkier BA, Kliebenstein DJ (2007) A Systems Biology Approach Identifies a R2R3 MYB Gene Subfamily with Distinct and Overlapping Functions in Regulation of Aliphatic Glucosinolates. Plos One **2**

Tamagnone L, Merida A, Stacey N, Plaskitt K, Parr A, Chang CF, Lynn D, Dow JM, Roberts K, Martin C (1998) Inhibition of phenolic acid metabolism results in precocious cell death and altered cell morphology in leaves of transgenic tobacco plants. Plant Cell **10:** 1801-1816

Tohge T, Alseekh S, Fernie A (2014) On the regulation and function of secondary metabolism during fruit development and ripening. J Exp Bot. **[Epub ahead of print]**

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

Tohge T, Nishiyama Y, Hirai MY, Yano M, Nakajima J, Awazuhara M, Inoue E, Takahashi H, Goodenowe DB, Kitayama M, Noji M, Yamazaki M, Saito K (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, Watanabe M, Hoefgen R, Fernie AR (2013) The evolution of phenylpropanoid metabolism in the green lineage. Critical Reviews in Biochemistry and Molecular Biology **48:** 123-152

Tohge T, Watanabe M, Hoefgen R, Fernie AR (2013) Shikimate and phenylalanine biosynthesis in the green lineage. Frontiers in plant science **4:** 62-62

Urbanczyk-Wochniak E, Usadel B, Thimm O, Nunes-Nesi A, Carrari F, Davy M, Blasing O, Kowalczyk M, Weicht D, Polinceusz A, Meyer S, Stitt M, Fernie

AR (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 Molecular Biology **60:** 773-792

Vanholme R, Storme V, Vanholme B, Sundin L, Christensen JH, Goeminne G, Halpin C, Rohde A, Morreel K, Boerjan W (2012) A Systems Biology View of Responses to Lignin Biosynthesis Perturbations in Arabidopsis. Plant Cell **24:** 3506-3529

Willits MG, Kramer CM, Prata RTN, De Luca V, Potter BG, Steffens JC, Graser G (2005) Utilization of the genetic resources of wild species to create a nontransgenic high flavonoid tomato. Journal of Agricultural and Food Chemistry **53:** 1231-1236

Winkel-Shirley B (2001) Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiol **126:** 485-493

Wu A, Allu AD, Garapati P, Siddiqui H, Dortay H, Zanor M-I, Asensi-Fabado MA, Munne-Bosch S, Antonio C, Tohge T, Fernie AR, Kaufmann K, Xue G-P, Mueller-Roeber B, Balazadeh S (2012) JUNGBRUNNEN1, a Reactive Oxygen Species-Responsive NAC Transcription Factor, Regulates Longevity in Arabidopsis. Plant Cell **24:** 482-506

Yonekura-Sakakibara K, Fukushima A, Nakabayashi R, Hanada K, Matsuda F, Sugawara S, Inoue E, Kuromori T, Ito T, Shinozaki K, Wangwattana B, Yamazaki M, Saito K (2012) Two glycosyltransferases involved in anthocyanin modification delineated by transcriptome independent component analysis in Arabidopsis thaliana. Plant Journal **69:** 154-167

Zhang Y, Butelli E, De Stefano R, Schoonbeek H-j, Magusin A, Pagliarani C, Wellner N, Hill L, Orzaez D, Granell A, Jones JDG, Martin C (2013) Anthocyanins Double the Shelf Life of Tomatoes by Delaying Overripening and Reducing Susceptibility to Gray Mold. Current Biology **23:** 1094-1100

Zhang Y, Butelli E, Martin C (2014) Engineering anthocyanin biosynthesis in plants. Current Opinion in Plant Biology **19:** 81-90

Zhao Q, Gallego-Giraldo L, Wang H, Zeng Y, Ding S-Y, Chen F, Dixon RA (2010) An NAC transcription factor orchestrates multiple features of cell wall development in Medicago truncatula. Plant Journal **63:** 100-114

Table 1. GC/MS profiling of Microtom WT and *Del/Ros1* transgenic line in fruit tissues

	Whole fruit			Flesh			Peel		
	WT	*Del/Ros1*	FC	WT	*Del/Ros1*	FC	WT	*Del/Ros1*	FC
Amino acid related									
alanine	42 ± 14	16 ± 6	**0.37***	86 ± 14	30 ± 10	**0.35*****	181 ± 29	100 ± 22	**0.55****
arginine	27 ± 3	26 ± 17	**0.97**	22 ± 3	18 ± 5	**0.85**	10 ± 2	11 ± 1	**1.17**
asparagine	5598 ± 1144	6073 ± 3162	**1.08**	5840 ± 214	5544 ± 1502	**0.95**	3673 ± 1126	3754 ± 209	**1.02**
aspartate	10516 ± 1414	9909 ± 928	**0.94**	9058 ± 2481	7851 ± 2002	**0.87**	8143 ± 1180	9515 ± 623	**1.17**
β-alanine	31 ± 10	26 ± 6	**0.85**	32 ± 6	9 ± 3	**0.29*****	23 ± 3	12 ± 1	**0.53*****
cystein	15 ± 5	14 ± 2	**0.91**	14 ± 3	8 ± 2	**0.59***	8 ± 1	10 ± 2	**1.22**
GABA	15034 ± 2309	12922 ± 4913	**0.86**	10913 ± 557	6622 ± 1464	**0.61*****	4332 ± 1614	5116 ± 431	**1.18**
glutamate	5141 ± 782	5698 ± 1107	**1.11**	4790 ± 447	4147 ± 491	**0.87**	6360 ± 2046	4211 ± 444	**0.66**
glutamine	1057 ± 454	1107 ± 130	**1.05**	507 ± 177	542 ± 99	**1.07**	840 ± 320	672 ± 61	**0.80**
glycine	301 ± 125	350 ± 296	**1.16**	347 ± 52	221 ± 25	**0.64****	446 ± 88	350 ± 42	**0.78**
glycolate	5 ± 3	7 ± 6	**1.47**	9 ± 2	7 ± 2	**0.78**	14 ± 3	14 ± 8	**1.05**
histidine	1166 ± 325	1081 ± 121	**0.93**	1238 ± 33	981 ± 189	**0.79***	784 ± 186	809 ± 92	**1.03**
isoleucine	780 ± 286	460 ± 74	**0.59**	796 ± 152	428 ± 103	**0.54****	852 ± 194	605 ± 29	**0.71***
luecine	933 ± 155	715 ± 134	**0.77**	968 ± 60	675 ± 153	**0.70***	840 ± 139	837 ± 23	**1.00**
lysine	1353 ± 109	1360 ± 204	**1.01**	1457 ± 224	1172 ± 228	**0.80**	650 ± 134	898 ± 80	**1.38***
methionine	321 ± 98	426 ± 143	**1.33**	584 ± 125	518 ± 113	**0.89**	251 ± 26	353 ± 60	**1.41***
ornithine	174 ± 33	112 ± 8	**0.64***	144 ± 22	120 ± 39	**0.83**	61 ± 21	84 ± 13	**1.37**
phenylalanine	792 ± 162	243 ± 69	**0.31*****	614 ± 108	312 ± 94	**0.51***	403 ± 110	380 ± 64	**0.94**
proline	1116 ± 790	1840 ± 1572	**1.65**	558 ± 214	538 ± 136	**0.96**	1003 ± 253	1031 ± 104	**1.03**
pyroglutamate	31449 ± 6598	28848 ± 5736	**0.92**	30286 ± 2671	26636 ± 2985	**0.88**	30722 ± 3812	27630 ± 2241	**0.90**
putrescine	3969 ± 1534	4470 ± 1524	**1.13**	3279 ± 345	3502 ± 650	**1.07**	1815 ± 253	2779 ± 294	**1.53****
serine	1220 ± 390	1212 ± 659	**0.99**	1182 ± 57	722 ± 194	**0.61****	1367 ± 205	1289 ± 40	**0.94**
threonate	189 ± 18	188 ± 8	**0.99**	192 ± 0	174 ± 26	**0.91**	145 ± 22	186 ± 11	**1.29***
threonine	565 ± 106	453 ± 85	**0.80**	770 ± 107	430 ± 122	**0.56****	535 ± 67	447 ± 29	**0.84**
tryptophan	825 ± 47	705 ± 84	**0.85***	893 ± 202	501 ± 101	**0.56***	848 ± 201	987 ± 81	**1.16**
tyrosine	349 ± 77	293 ± 119	**0.84**	719 ± 133	219 ± 45	**0.30*****	1063 ± 367	381 ± 66	**0.36***
valine	580 ± 155	472 ± 44	**0.81**	627 ± 79	505 ± 153	**0.81**	769 ± 143	769 ± 58	**1.00**
Organic acids									
2OXglutarate	28 ± 15	39 ± 16	**1.42**	13 ± 3	20 ± 5	**1.62***	21 ± 5	35 ± 6	**1.67***
ascorbate	5 ± 2	4 ± 1	**0.87**	5 ± 2	3 ± 0	**0.68**	7 ± 3	4 ± 1	**0.66**
benzoate	16 ± 7	14 ± 5	**0.89**	20 ± 4	13 ± 3	**0.67***	19 ± 6	16 ± 1	**0.80**
citrate	9116 ± 1735	8889 ± 2403	**0.98**	7166 ± 1304	6291 ± 1144	**0.88**	6244 ± 1297	6876 ± 419	**1.10**
fumarate	14 ± 4	16 ± 4	**1.13**	13 ± 1	14 ± 2	**1.14**	28 ± 4	29 ± 1	**1.06**
glycerate	2 ± 2	2 ± 1	**0.99**	5 ± 2	5 ± 2	**1.13**	6 ± 1	6 ± 2	**0.93**
malate	1925 ± 506	2326 ± 858	**1.21**	1868 ± 394	1776 ± 387	**0.95**	1367 ± 128	1675 ± 183	**1.23***
nicotinate	208 ± 48	185 ± 90	**0.89**	256 ± 11	188 ± 21	**0.73****	199 ± 22	162 ± 9	**0.81***
phosphorate	7875 ± 1778	8948 ± 1558	**1.14**	8272 ± 769	7202 ± 1180	**0.87**	11443 ± 1512	12606 ± 757	**1.10**
putative citramalate	7 ± 2	8 ± 6	**1.22**	17 ± 5	7 ± 2	**0.41***	16 ± 2	8 ± 2	**0.49*****
putative galactarate	73 ± 22	289 ± 98	**3.96***	54 ± 4	192 ± 54	**3.57****	80 ± 9	442 ± 44	**5.50*****
putative galacturonate	68 ± 26	58 ± 10	**0.85**	45 ± 9	36 ± 9	**0.79**	52 ± 12	59 ± 12	**1.14**
putative glatactonate	400 ± 138	712 ± 281	**1.78**	497 ± 77	946 ± 199	**1.90****	805 ± 64	1230 ± 56	**1.53*****
quinate	264 ± 86	502 ± 125	**1.90***	134 ± 10	407 ± 114	**3.04****	170 ± 31	514 ± 51	**3.03*****
succinate	3 ± 2	3 ± 1	**0.88**	8 ± 5	4 ± 1	**0.46**	11 ± 5	9 ± 3	**0.86**
Sugar related									
1,6-anhydroglucose	116 ± 37	128 ± 23	**1.11**	101 ± 14	87 ± 13	**0.87**	116 ± 25	121 ± 16	**1.04**

fructose	334 ± 103	322 ± 82	**0.97**	278 ± 47	280 ± 77	**1.00**	124 ± 43	180 ± 11	**1.45***
maltose	25 ± 11	29 ± 21	**1.16**	22 ± 1	14 ± 2	**0.65****	34 ± 9	28 ± 3	**0.83**
galactose	6948 ± 1363	1529 ± 1357	**0.22****	8069 ± 482	1931 ± 419	**0.24*****	6887 ± 1495	1790 ± 282	**0.26*****
gluconate-6-phosphate	25 ± 6	30 ± 10	**1.22**	21 ± 3	20 ± 2	**0.93**	23 ± 4	26 ± 4	**1.11**
glucose	6698 ± 3036	6763 ± 2006	**1.01**	4709 ± 1131	4859 ± 1089	**1.03**	1652 ± 684	2767 ± 170	**1.68***
glucose-6-phosphate	57 ± 11	70 ± 15	**1.22**	48 ± 4	48 ± 5	**1.00**	58 ± 13	73 ± 11	**1.26**
sucrose	1457 ± 286	1625 ± 359	**1.11**	338 ± 142	459 ± 141	**1.36**	676 ± 73	825 ± 81	**1.22***
Others									
3CGAcis	6 ± 2	18 ± 12	**3.16**	2 ± 0	9 ± 3	**3.80****	20 ± 4	20 ± 1	**0.99**
3CGAtrans	9 ± 5	85 ± 70	**9.21**	1 ± 0	28 ± 11	**21.60****	18 ± 9	73 ± 5	**3.99*****
4CGAtrans	12 ± 9	78 ± 33	**6.42****	4 ± 1	55 ± 22	**15.35****	15 ± 4	108 ± 5	**7.28*****
5CGAtrans	8 ± 6	38 ± 30	**4.93**	4 ± 2	23 ± 9	**5.64****	9 ± 2	29 ± 2	**3.11*****
5,6-dihydrouracil	11462 ± 2836	10783 ± 2359	**0.94**	11499 ± 443	9683 ± 1374	**0.84***	11790 ± 2136	10322 ± 1332	**0.88**
adenine	42 ± 8	46 ± 11	**1.08**	57 ± 6	44 ± 6	**0.77***	54 ± 12	42 ± 4	**0.77**
dehydroascorbate dimer	2193 ± 850	1780 ± 345	**0.81**	1578 ± 109	1268 ± 255	**0.80**	1670 ± 265	1480 ± 234	**0.89**
galactinol	126 ± 112	179 ± 64	**1.42**	71 ± 84	21 ± 10	**0.30**	119 ± 63	92 ± 23	**0.77**
glycerol	87 ± 22	79 ± 26	**0.91**	231 ± 70	193 ± 19	**0.84**	284 ± 50	218 ± 16	**0.77***
glycerol-3-phosphate	21 ± 6	23 ± 7	**1.08**	40 ± 22	18 ± 2	**0.44**	42 ± 22	38 ± 11	**0.89**
guanosine	36 ± 13	37 ± 9	**1.05**	42 ± 4	36 ± 7	**0.86**	65 ± 9	57 ± 4	**0.88**
myo-inositol	1359 ± 730	1742 ± 965	**1.28**	974 ± 351	942 ± 41	**0.97**	1414 ± 412	1831 ± 57	**1.30**
nicotianamide	50 ± 15	53 ± 12	**1.07**	49 ± 4	48 ± 8	**0.97**	61 ± 6	73 ± 8	**1.20**
threitol	67 ± 19	39 ± 9	**0.58***	82 ± 10	38 ± 9	**0.47*****	86 ± 8	52 ± 3	**0.61*****
uracil	26 ± 6	24 ± 5	**0.90**	35 ± 3	33 ± 7	**0.96**	50 ± 7	41 ± 2	**0.82**
urea	41 ± 19	44 ± 32	**1.07**	93 ± 26	54 ± 8	**0.58***	131 ± 53	146 ± 35	**1.12**

*,P<0.05; **,P<0.01; ***, P<0.001

Table 2. Upregulated genes in *Del/Ros* tomato

				WT		*Del/Ros*					
ITAG2.3	Ara orthologs	Gene name	function	B+1w count	B+4w count	B+1w count	B+4w count	**B+1w FC**	**B+4w FC**		
Flavonoid biosynthetic genes											
Solyc05g010320	At3g55120	CHI	chalcone isomerase	1	1	13	10	**12**	**11**	*ant1*	*pap1*-D
Solyc11g066580	At5g07990	F3'5'H	Flavonoid-3'-monooxygenase	1	1	1052	485	**940**	**541**		*pap1*-D
Solyc08g080040	At4g22880	ANS	leucoanthocyanidin dioxygenase	1	1	1241	243	**1108**	**181**		*pap1*-D
Solyc02g085020	At5g42800	DFR	Dihydrokaempferol-4-reductase	1	1	898	217	**802**	**242**		*pap1*-D
Solyc09g082660	At1g67980	AnthOMT	anthocyanin-*O*-methyltransferase	1	1	228	112	**204**	**125**		
Solyc10g083440	At5g17050	A3GT	*anthocyanin-3-*O*-glucosyltransferase	52	7	1960	575	**38**	**80**	*ant1*	*pap1*-D
Solyc09g059170	At5g54010	A3G2"GT	*anthocyanin-3-*O*-glc-2"-*O*-glucosyltransferase	12	1	1628	130	**135**	**145**		
Solyc12g098590	At4g14090	A5GT	*anthocyanin-5-*O*-glucosyltransferase	5	1	4978	1291	**926**	**1441**	*ant1*	*pap1*-D
Solyc12g088170	At3g26040	AAT	*anthocyanin acyltransferase	1	1	1764	837	**1576**	**935**		
Solyc02g081340	At5g17220	GST	glutathione-*S*-transferase	1	1	4578	1630	**4088**	**1820**	*ant1*	*pap1*-D
Solyc03g025190	At4g25640	FFT	flower flavonoid transporter	1	1	1003	390	**896**	**435**	*ant1*	
Shikimate biosynthetic gene											
Solyc02g088460	At3g29200	CM1	chorismate mutase 1	12	6	232	55	**19**	**9**		
Phenylpropanoid biosynthetic gene											
Solyc09g007920	At2g37040	SlPAL4	phenylalanine ammonia-lyase	8	3	77	42	**10**	**16**		
Solyc05g056170	At3g53260	SlPAL6	phenylalanine ammonia-lyase	4	2	75	38	**18**	**17**		
Solyc01g079240	At2g47240	Sl4CL6	4-coumarate-CoA ligase	10	3	245	51	**25**	**16**		
Solyc02g086770	At5g14700	CCR	cinnamoyl-CoA reductase	2	1	70	15	**39**	**16**		
Others											
Solyc09g065100	At4g09820	bHLH		2	1	312	138	**199**	**154**		*pap1*-D
Solyc03g120620	At1g79840	MYB		5	1	91	43	**17**	**32**		
Solyc07g007750	At2g02130	PDF	plant defensin	98	29	13553	8289	**139**	**285**		
Solyc12g099260	At5g49460	ACLB-2	succinyl-CoA synthetase subunit	1	1	164	9	**146**	**10**		
Solyc07g061800	At1g17100	HBP1	SOUL heme-binding family protein	3	3	312	23	**93**	**8**		
Solyc01g005800	At3g13600		calmodulin-binding protein	2	2	80	18	**51**	**8**		
Solyc07g062490	At1g78830	Curculin	curculin-like lectin family protein	4	2	181	22	**50**	**12**		
Solyc06g083080	At5g48500		light responsive unknown	2	1	72	14	**46**	**16**		
Solyc04g078140	At5g48810	ATB5-B	cytochrome b5 B	17	16	746	216	**44**	**14**		
Solyc01g058320	At3g17120		unknown protein	1	2	37	18	**33**	**10**		
Solyc09g010690	At2g28840	XBAT31	ankyrin repeat family protein	2	1	40	9	**25**	**10**		
Solyc06g050870	At3g05550		hypoxia induced protein	1	1	15	24	**14**	**18**		
Solyc08g081790	At4g23690	DIR6	disease resistance-responsive family protein	6	1	82	10	**14**	**11**		

FC, fold change, "*pap-D*" indicates upregulated genes listed in *pap-D* Arabidopsis leaf (Tohge *et al*., 2005). "*ant1*" indicated upregulated genes found in ANT1 transgenic tomato seedling (Mathews *et al*., 2003). *: gene function annotated in this study.

Table 3. Anthocyanin-3-*O*-(4'''-phenylacyl-rutinoside)-5-*O*-glucosides and flavonol-3-*O*-(4'''-phenylacyl)-rutinosides found in plant species

Flavonoid	(total number of phenylacylated flavonoid)	Formula	Species	Plant name
anthocyanins				
delphinidin derivatives (37)				
C00006893	delphinidin-3-(4'''-*trans*-*p*-coumaroyl-Rut)-5-Glc	$C_{42}H_{47}O_{23}$	Iridaceae	*Iris tingitana*
	Synonym: violanin, nasunin		Solanaceae	*Petunia x hybrida*
			Solanaceae	*Solanum melongena*
			Solanaceae	*Solanum tuberosum*
			Solanaceae	*Capsicum annuum*
			Violaceae	*Viola tricolor*
C00014796	delphinidin-3-(4'''-(4''''-Glc)-*trans*-*p*-coumaroyl)-Rut-5-Glc	$C_{48}H_{57}O_{28}$	Solanaceae	*Petunia reitzii*
C00014797	Delphinidin-3-(4'''-*trans*-caffeoyl)-Rut-5-Glc	$C_{42}H_{47}O_{24}$	Solanaceae	*Petunia occidentalis*
C00014798	Delphinidin-3-(4'''-*trans*-*p*-coumaryl)-Rut-5-Glc	$C_{42}H_{47}O_{23}$	Solanaceae	*Solanum melongena*
petunidin derivatives (9)				
C00006900	petunidin-3-(4'''-*trans*-*p*-coumaroyl-Rut)-5-Glc	$C_{43}H_{49}O_{23}$	Iridaceae	*Iris spp.*
	Synonym: petanin		Solanaceae	*Petunia x hybrida*
			Solanaceae	*Solanum nigrum*
			Solanaceae	*Solanum tuberosum*
C00011099	petunidin-3-(4'''-(6''''-O-caffeoyl-Glc)-*p*-coumaroyl-Rut)-5-Glc	$C_{58}H_{65}O_{31}$	Solanaceae	*Petunia x hybrida*
C00014855	petunidin-3-(4'''-caffeoyl-Rut)-5-Glc	$C_{43}H_{49}O_{24}$	Solanaceae	*Solanum tuberosum*
C00014861	petunidin-3-(4'''-(4''''-Glc)-*p*-coumaroyl-Rut)-5-Glc	$C_{49}H_{59}O_{28}$	Solanaceae	*Petunia occidentalis*
C00014862	petunidin-3-(4'''-feruroyl-Rut)-5-Glc	$C_{44}H_{51}O_{24}$	Solanaceae	*Solanum tuberosum*
malvidin derivatives (22)				
C00006914	malvidin-3-(4'''-caffeoyl-Rut)-5-Glc	$C_{44}H_{51}O_{24}$	Solanaceae	*Petunia x hybrida*
C00011074	malvidin-3-(4'''-(6''''-caffeoyl-Glc)- p-coumaroyl-Rut)-5-Glc	$C_{59}H_{67}O_{31}$	Solanaceae	*Petunia guarapuavensis*
			Solanaceae	*Petunia hybrida cultivar*
C00011075	malvidin-3-(4'''- (6''''-caffeoyl-Glc)- caffeoyl-Rut)-5-Glc	$C_{59}H_{67}O_{32}$	Solanaceae	*Petunia guarapuavensis*
			Solanaceae	*Petunia hybrida cultivar*
C00011100	malvidin-3-(4'''- (6''''-feruroyl-Glc)- *p*-coumaroyl-Rut)-5-Glc	$C_{60}H_{69}O_{31}$	Solanaceae	*Petunia x hybrida*
C00011101	malvidin-3-(4'''-(6''''-Glc-*p*-coumaroyl)-*p*-coumaroyl-Rut)-5-Glc	$C_{59}H_{67}O_{30}$	Solanaceae	*Petunia x hybrida*
C00011102	malvidin-3-(4'''-*p*-coumaroyl-Rut)-5-Glc	$C_{44}H_{51}O_{23}$	Solanaceae	*Petunia x hybrida*
C00014829	malvidin-3-(4'''-feruroyl-Rut)-5-Glc	$C_{45}H_{53}O_{24}$	Solanaceae	*Solanum tuberosum*
peonidin derivatives (10)				
C00006873	peonidin-3-(4'''-*p*-coumaroyl-Rut)-5-Glc	$C_{43}H_{49}O_{22}$	Solanaceae	*Petunia x hybrida*
	Synonym: peonanin		Solanaceae	*Solanum nigrum*
			Solanaceae	*Solanum phureja*
			Solanaceae	*Solanum tuberosum*
C00006874	peonidin-3-(4'''-caffeoyl-Rut)-5-Glc	$C_{43}H_{49}O_{23}$	Convolvulaceae	*Ipomoea indivisa*
			Solanaceae	*Petunia x hybrida*
			Solanaceae	*Solanum tuberosum*
pelargonidin derivatives (41)				
C00006789	pelargonidin-3-(4'''-*p*-coumaryl-Rut)-5-Glc	$C_{42}H_{47}O_{21}$	Caryophyllaceae	*Silene dioica*
	Synonym: pelanin		Solanaceae	*S.andigenaxS.tuberosum*
C00014853	pelargonidin-3-(4'''-feruloyl-Rut)-5-Glc	$C_{43}H_{49}O_{22}$	Solanaceae	*S.andigenaxS.tuberosum*
cyanidin derivatives (73)				
C00006856	cyanidin-3-(4'''-caffeyl-Rut)-5-Glc	$C_{42}H_{47}O_{23}$	Caryophyllaceae	*Silene dioica*
68156-54-7	cyanidin-3-(4'''-caffeyl-Rut)	$C_{36}H_{37}O_{18}$	Caryophyllaceae	*Silene dioica*
flavonols				
kaempferol derivatives (83)				
C00005905	kaempferol-3-(4'''-(3'''-Rha)-*p*-coumaroyl-Rut	$C_{42}H_{46}O_{21}$	Gleicheniaceae	Dicranopteris linearis
quercetin derivatives (42)				
no entry				
isorhamnetin derivatives (10)				
no entry				
myricetin derivatives (3)				
no entry				

Flavonoids were searched by KNaPSAcK database and The handbook of natural flavonoids edited by Horbone and Baxter (1999). References are applied in Supplemental File S3.

Table 4. Enzyme activity of recombinant SlFdAT in *E.coli.*

	nkat/mg	Relative Activity (%)[a]
Acyl acceptor[b]		
cyanidin-3-rutinoside	0.046 ± 0.012	100.0
delphinidin-3-rutinoside	0.090 ± 0.036	194.4
quercetin-3-rutinoside	0.052 ± 0.012	112.4
kaempferol-3-rutinoside	0.019 ± 0.001	42.1
Acyl-donor[c]		
p-coumaroyl-CoA	0.046 ± 0.012	100.0
feruloyl-CoA	0.041 ± 0.007	88.6
caffeoyl-CoA	0.018 ± 0.001	40.0
cinnamoyl-CoA	0.034 ± 0.015	74.8
sinapoyl-CoA	0.004± 0.004	8.0
malonyl-CoA	0	0.0

[a] Relative activity was calculated by comparison to the activity of the enzyme with cyanidin 3-rutinoside as acyl acceptor and *p*-coumaroyl-CoA as acyl-donor. [b] The reactions were carried out using *p*-coumaroyl CoA as the acyl-donor. [c] The reactions were carried out using cyanidin 3-rutinoside as the acyl acceptor

FIGURE LEGENDS

Figure 1. Comparative metabolite profiling between WT and *Del/Ros1* purple tomatoes

The whole fruits of ripe fruit harvested at a time point of Breaker +1 week (B+1w) were used for the analysis. A) LC/MS total ion chromatogram (TIC, negative ion detection) of WT and *Del/Ros1* tomatoes. B) Chemical structure of major flavonoid aglycones characterized previously. C) Changes in major phenolics in *Del/Ros1* tomatoes compared to WT. Intensity of log-scaled fold changes is indicated by colour; red, increased in *Del/Ros1* tomatoes; blue, decreased in *Del/Ros1* tomatoes both compared to WT. Descriptions and abbreviation of the metabolites (1-27) are given in Supplemental Table 1. Heatmap shows values displayed on a $\log_2$ scale (-3 to 3) for fold change (Log_2FC) compared to WT.

Figure 2. Gene expression profiles in *Del/Ros1* tomatoes

Gene expression profiling was performed by RNA-seq for two time points: ripe stage (Breaker +1 week, B+1w) and over-ripe stage (Breaker + 4 weeks, B+4w). A) Expression level of phenylpropanoid biosynthetic genes characterized or annotated previously. B) Venn diagram of a) up- (>8.0 fold) or b) down- (<0.125 fold) regulated genes at both developmental stages. Intensity of log-scaled fold change (-3 to 3) was visualized using colour; red, up-regulated in *Del/Ros1* tomatoes; blue, down-regulated in *Del/Ros1* tomatoes.

Figure 3. Characterization of major anthocyanins in tomato

A) Tomato anthocyanins characterized in this study. B) Co-elution profiles of purple tomato extracts for peak characterization of endogenous anthocyanins in domesticated tomato (*S. lycopersicum*, M82) and wild tomato (*S. pennellii*) in leaves. LC/MS ion extracted chromatogram (TA1, petanin, 933 *m/z*; TA2, nasunin 919 *m/z* in positive ion detection) are shown.

Figure 4. Phylogenetic analysis of candidate genes belonging to UGT and BAHD genes involved in flavonoid biosynthesis

Molecular phylogenetic tree of the amino acid sequences of the A) flavonoid glycosyltransferases and B) flavonoid acyltransferases from tomato. The amino acid

sequences were aligned using the multiple sequence alignment MEGA 5.1 (http://www.megasoftware.net/). C) A total of 31 plant species were used for orthologue gene cluster (ORTH) analysis using plaza dicot v3.0 (http://bioinformatics.psb.ugent.be/plaza//). Effects are indicated by arrows conferred by colour; red, up-regulated in *Del/Ros1* tomatoes. D) Molecular phylogenetic tree of the amino acid sequences corresponding to the ORF encoding orthologous to SlFdAT1 (Solyc12g088170). The amino acid sequences were aligned using the multiple sequence alignment MEGA 5.1 (http://www.megasoftware.net/).

Figure 5. Functional characterization of SlFdAT1

A)-D) HPLC chromatograms for assays of the SlFdAT acyltransferase from *S. lycopersicum.* A) Activity of recombinant protein expressed in *E.coli* with cyanidin-3-*O*-rutinoside as acyl-acceptor and *p*-coumaroyl-CoA as acyl-donor, B) delphinidin-3-*O*-rutinoside as acyl-acceptor and *p*-coumaroyl-CoA as acyl-donor, C) quercetin-3-*O*-rutinoside as acyl-acceptor and *p*-coumaroyl-CoA as acyl-donor, D) kaempferol-3-*O*-rutinoside as acyl-acceptor and *p*-coumaroyl-CoA as acyl-donor. E) and F) HPLC chromatograms of the flower extracts of WT and the *35S:SlFdAT1* tobacco lines showing common peaks, peak1, cyanidin-3-*O*-rutinoside; peak 2, pelargonidin-3-*O*-rutinoside, and new peaks, peak 3, cyanidin-3-*O*-(*p*-coumaroyl)-rutinoside; peak 4, pelargonidin-3-*O*-(*p*-coumaroyl)-rutinoside; peak 5, quercetin-3-*O*-(*p*-coumaroyl)-rutinoside.

Figure 6. Summary of differences in regulation of anthocyanin biosynthesis between tomato and Arabidopsis.

Effects are indicated by arrows shown by colour; red; up-regulated in A) *Del/Ros1* tomato or B) *pap1*-D Arabidopsis.

Figure 1 Tohge *et al.*

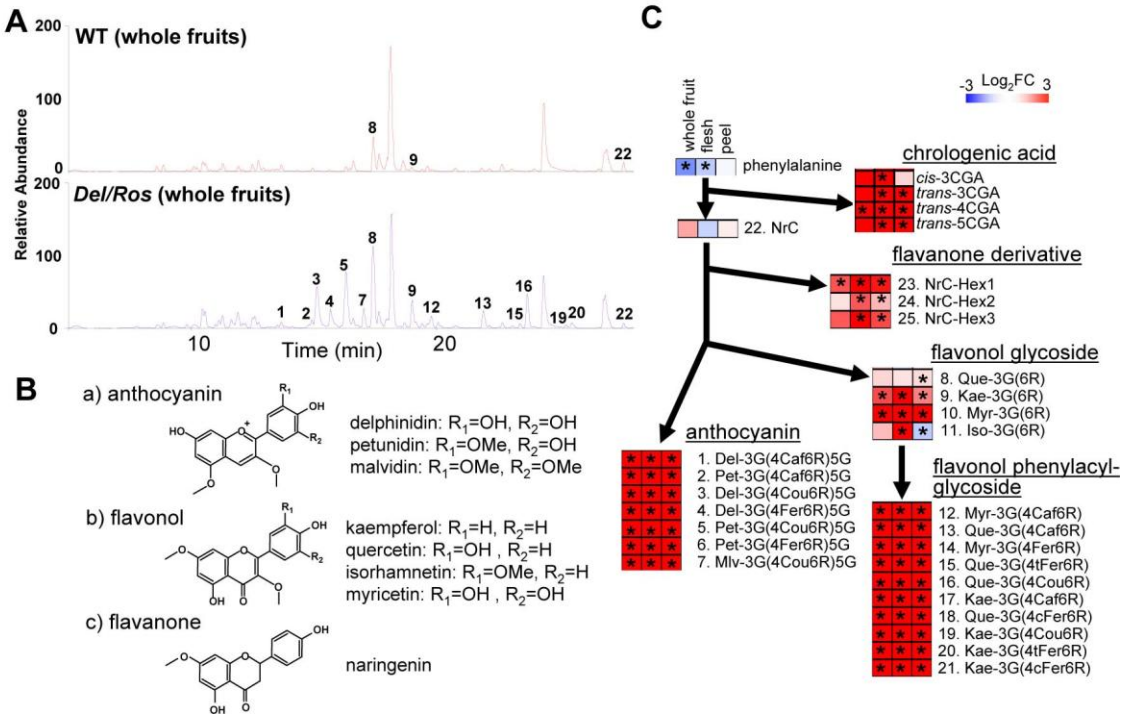

Figure 2 Tohge *et al.*

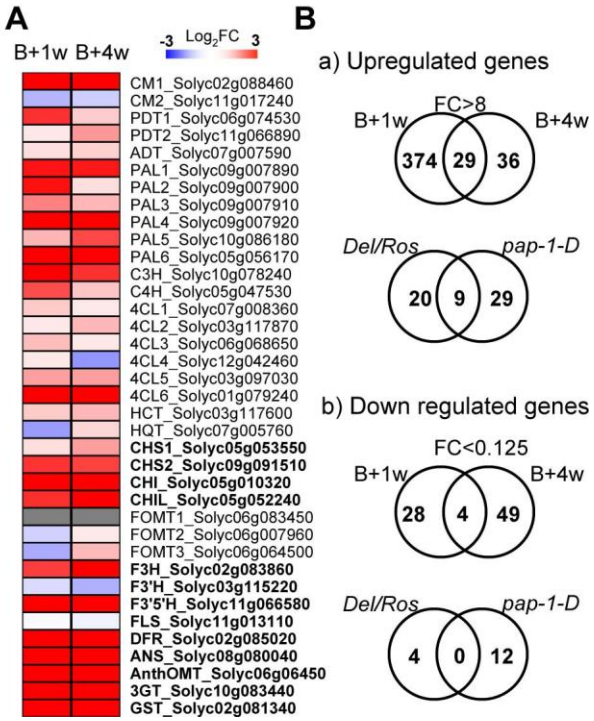

Figure 3 Tohge *et al.*

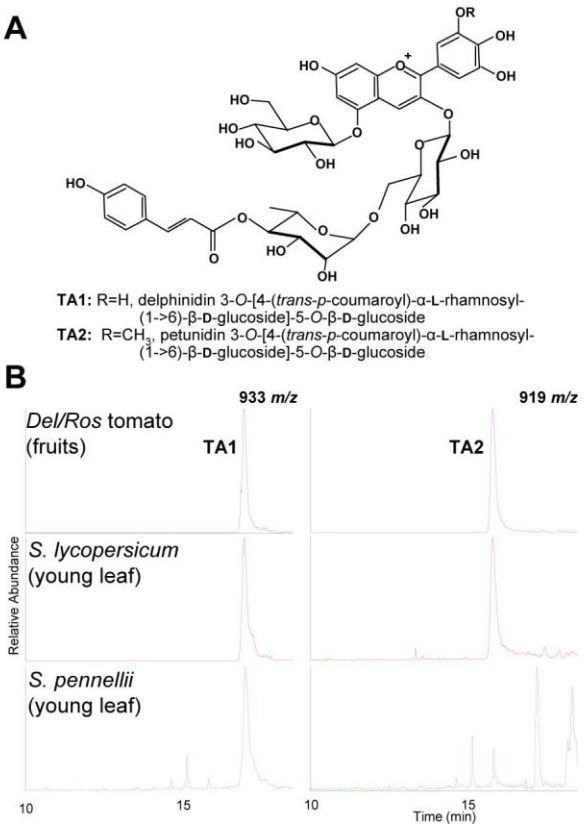

Figure 4 Tohge *et al.*

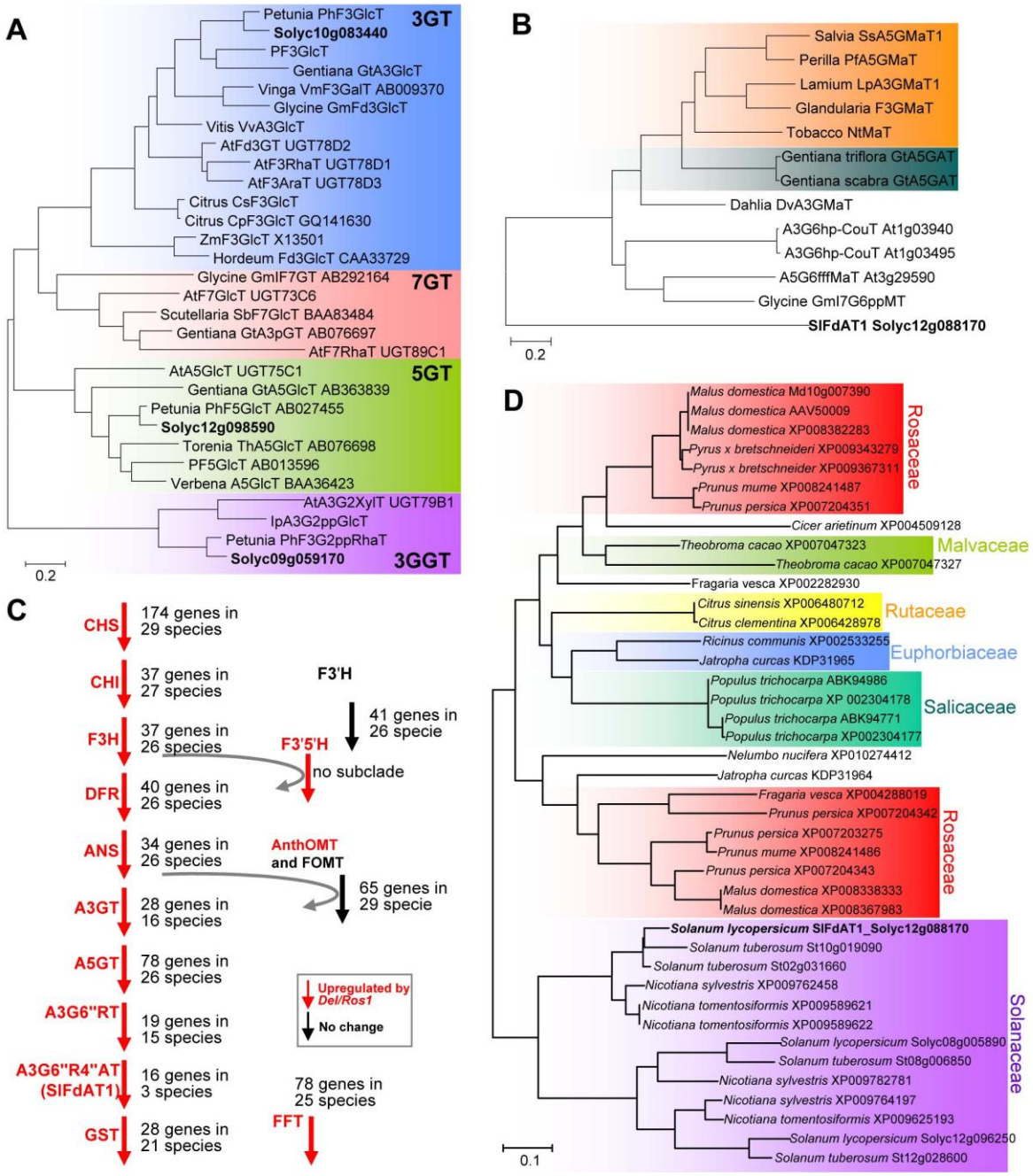

Figure 5 Tohge *et al.*

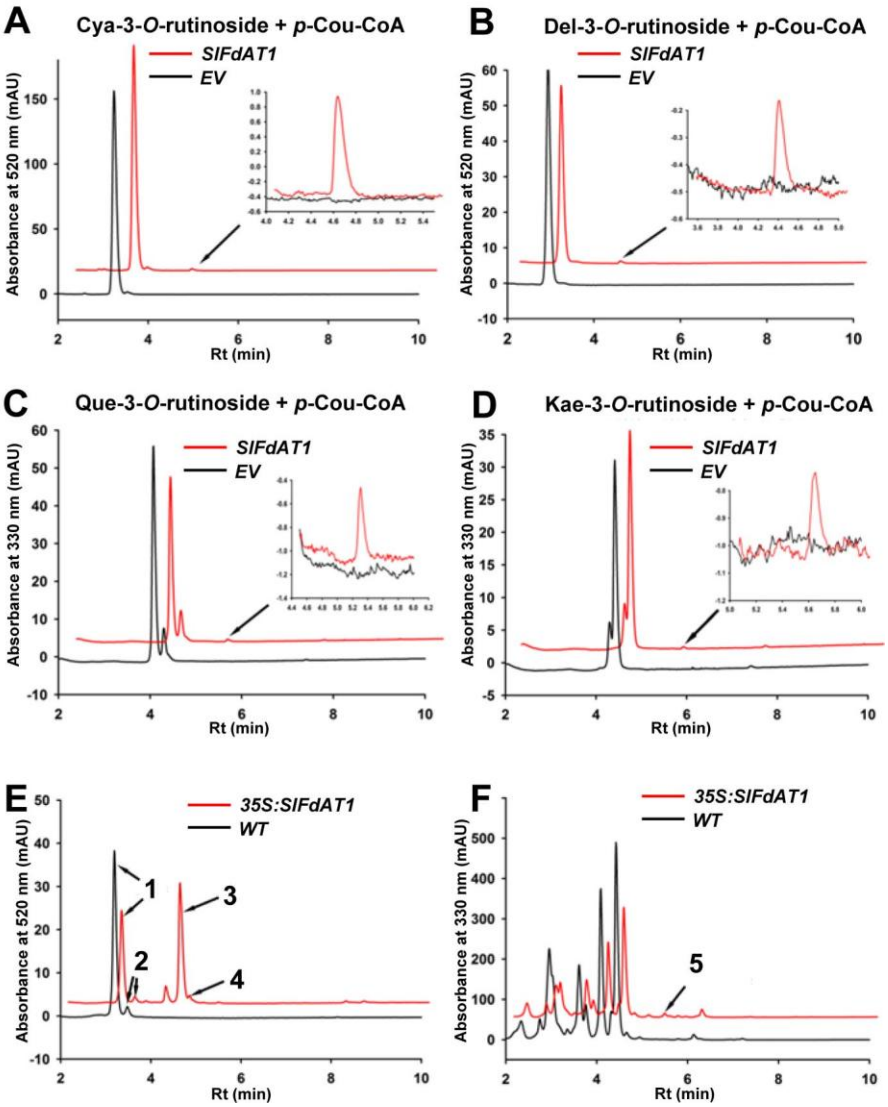

Figure 6 Tohge *et al.*