the plant journal

SEB Society for Experimental Biology

The Plant Journal (2015) **83,** 686–704 doi: 10.1111/tpj.12920

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

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

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

Received 19 March 2015; revised 15 June 2015; accepted 16 June 2015; published online 23 June 2015.
*For correspondence (e-mail fernie@mpimp-golm.mpg.de).

SUMMARY

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 (*Antirrhinum majus*), paying particular attention to changes in phenylpropanoids compared to controls. These analyses revealed multiple changes, including depletion of rutin and naringenin chalcone, and enhanced levels of anthocyanins and phenylacylated flavonol derivatives. We isolated and characterized the chemical structures of the two most abundant anthocyanins, which were shown by NMR 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 purple fruit and wild-type fruit, we obtained 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 anthocyanin biosynthetic genes, 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 over-expression 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 various plant species.

Keywords: anthocyanin metabolism, tomato, Solanaceous species, RNA-seq, acyltransferase, glycosyltransferase.

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), and defence against herbivores and pathogens (Brechenmacher *et al.*, 2010; Huang *et al.*, 2010). Their widespread distribution amongst seed plants (Tohge *et al.*, 2013a,b), 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 that are rich in phenolic compounds. The broad spectrum of biological properties of individual phenolics, their bioavailability and the range of levels in various foodstuffs are generally uncharacterized, and even the chemical identity of the individual compounds themselves has often not been established. Nevertheless, elucidation of these compounds and their biological properties provides an important foundation upon which strategies for metabolic engineering or biofortification may be based.

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 may 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 strategies to modify the biosynthesis of phenolics and hence alter the composition of tomato fruit flavonoids have been tested. The most common approach is altered expression of specific endogenous genes, including *ANT1*, encoding a Myb transcription factor, *AFT*, encoding a Myb transcription factor, and *DE-ETIOLATED1*, encoding a chromatin remodelling factor (Jones *et al.*, 2003; Mathews *et al.*, 2003; Enfissi *et al.*, 2010), and ectopic expression of chalcone isomerase from petunia (*Petunia* x *hybrida*) or the transcription factors *LC* and *C1* from maize (*Zea mays*) (Muir *et al.*, 2001; Bovy *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, with their accumulation being confined to the surface layers.

However, ectopic fruit-specific expression of the snapdragon (*Antirrhinum majus*) transcription factors Delila (*Del*, bHLH) and Rosea1 (*Ros1*, Myb) 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, and thus combined expression of the bHLH and Myb transcription factors overcame the limited bHLH expression in tomato fruit and resulted in induction of the gene encoding flavonoid-3′5′-hydroxylase (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 over-expression, as, 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 using 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 greatly improved since the proof-of-concept study that first took this approach, i.e. the combined transcriptomic and metabolomic evaluation of the AtMYB75 (AtPAP1) activation-tagged Arabidopsis line by 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 (Carrari *et al.*, 2006; Mounet *et al.*, 2009; Rohrmann *et al.*, 2011), more comprehensive analyses 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 with the wild-type control, as well as the performance of 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 characterized 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 aubergine/eggplant (Kuroda and Wada, 1933; Sakamura *et al.*, 1963) (*Solanum melongena*) and petunia (Schram *et al.*, 1983), 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 (anthocyanin-3–*O*-glucosyltransferase, SlA3GlcT), Solyc12g098590 (anthocyanin-5–*O*-glucosyltransferase, SlA5GlcT), Solyc09g059170 (anthocyanin-3–*O*-glucoside-6″–*O*-rhamnosyltransferase, Sl A3Glc6″RhaT) and Solyc12g088170 (anthocyanin-3–*O*-rutinoside-4‴–*O*-phenylacyltransferase). Further investigation of the function of Solyc12g088170 (SlFdAT1) using

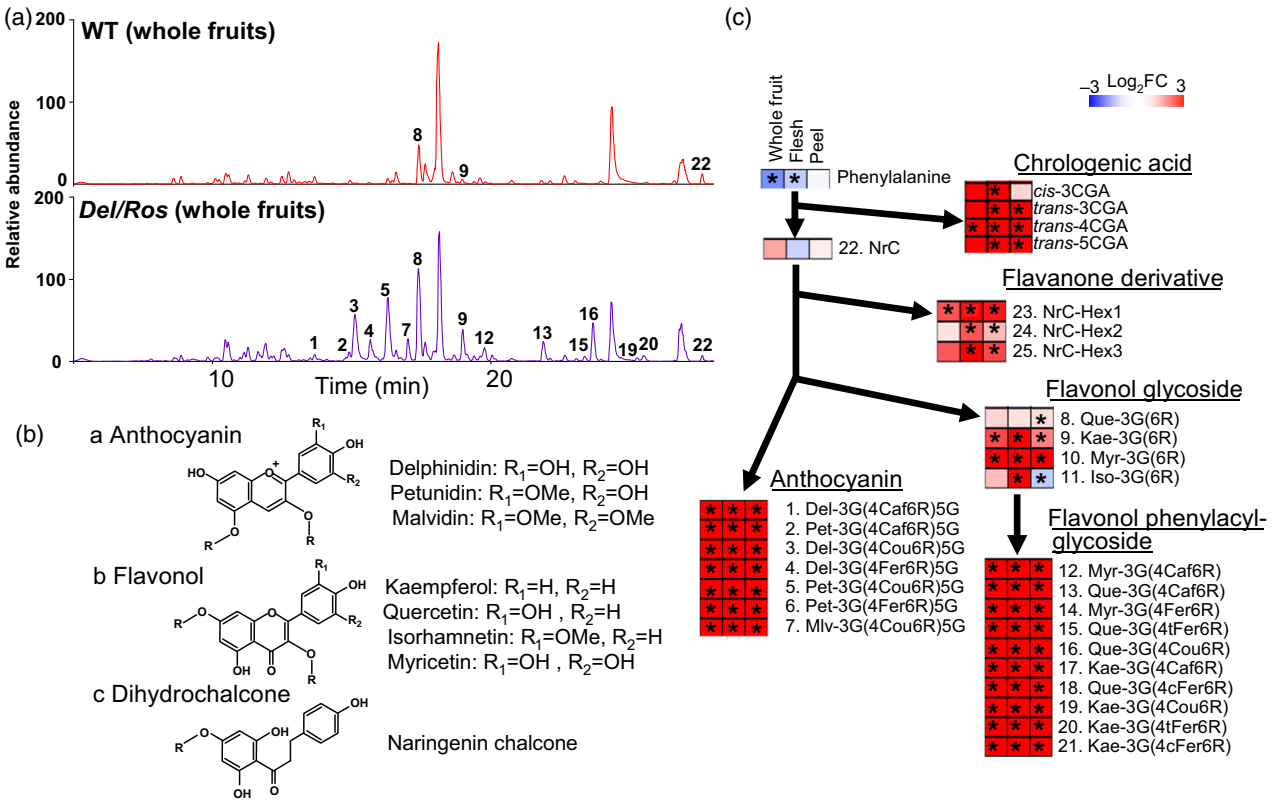


Figure 1. Comparative metabolite profiling between WT and *Del*/*Ros1* purple tomatoes.
Whole fruits of ripe fruit harvested at the time point breaker + 1 week (B + 1w) were used for the analysis.
(a) LC/MS total ion chromatogram (negative ion detection) for WT and *Del*/*Ros1* tomatoes.
(b) Chemical structure of major flavonoid aglycones characterized previously (Slimestad and Verheul, 2009).
(c) Changes in major phenolics in *Del*/*Ros1* tomatoes compared to WT. The intensity of log-scaled fold changes is indicated using colour: red, increased in *Del*/*Ros1* tomatoes; blue, decreased in *Del*/*Ros1* tomatoes, both compared to WT. Descriptions and abbreviations of the metabolites (1–27) are given in Table S1. The heatmaps show values displayed on a $\log_2$ scale (−3 to 3) for fold change compared to WT.

recombinant enzyme assays and metabolite profiling of transgenic tobacco (*Nicotiana tabacum*) 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 performed LC/MS-based secondary metabolite profiling of extracts from whole fruits, flesh and peel from wild-type and *Del*/*Ros1* purple tomatoes harvested at breaker + 1 week (B + 1w) (Butelli *et al.*, 2008). Representative chromatograms are shown in Figure 1(a) and Figure S1, in which it may clearly be seen that the *Del*/*Ros1* line has a richer diversity of secondary metabolites than the control for both whole-fruit and peel samples. A total of seven anthocyanins, four flavonols, ten phenylacylated flavonols and four naringenin derivatives were identified and annotated across both genotypes (Figure 1b). The anthocyanins and phenylacylated flavonols were essentially detected only in the *Del*/*Ros1* fruit, and large quantitative differences were seen for the flavonols and naringenin derivatives between the lines (Figure 1c and Table S1). Indeed, in 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 levels in the *Del*/*Ros1* fruit were only 60% of those observed in the WT control, although the levels of 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 in Figure S2, tomatine-related glycoalkaloids were twofold higher in *Del*/*Ros1* fruit, whilst the level of esculeoside-type glycoalkaloids was slightly lower (Figure S2 and Table S1). Given that reduction of tomatine-type glycoalkaloids is a strong indicator of fruit ripening (Schwahn *et al.*, 2014; Tohge *et al.*, 2014), the higher levels of tomatine-type glycoalkaloids in *Del*/*Ros1* fruit probably reflect a delay in ripening, confirming previous observations (Zhang *et al.*, 2013).

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

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

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 + 1 week (B + 1w) and breaker + 4 weeks (B + 4w) (Table S2). Changes in gene expression were evaluated on a global scale using the most recent update of the tomato MAPMAN files (Urbanczyk-Wochniak *et al.*, 2006) (http://mapman.gabipd.org/) (Figure S3). Genes associated with flavonoid, phenylpropanoid, cell wall biosynthesis, wax and aromatic amino acid biosynthesis, as well as lipid metabolism and histidine degradation, were up-regulated 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 down-regulated 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 in the purple fruit, as well as some associated with the tricarboxylic acid cycle (consistent with the elevated levels of 2OG and malate), sugar metabolism and sulfur assimilation, but levels of transcripts associated with Gln, Pro and Arg synthesis decreased, as were those for the branched chain amino acids (consistent with the observed changes in the levels of the metabolites themselves), and most enzymes involved in lipid metabolism were down-regulated. These changes in transcripts are consistent with a delay in ripening as previously observed (Zhang *et al.*, 2013). At the B + 4w stage, the only specific differences were down-regulation of transcripts associated with the Calvin–Benson cycle, the photosystems and the tricarboxylic acid cycle.

We next performed a targeted analysis of genes associated with anthocyanin metabolism. Figure 2(a) shows the gene expression of known or previously annotated genes involved in phenylalanine, phenylpropanoid and flavonoid biosynthesis in tomato. Using this approach, we were able to annotate a total of 37 genes comprising five genes involved in phenylalanine biosynthesis, including those encoding chorismate mutase, prephenate aminotransferase and arogenate dehydratase, 13 genes of general phenylpropanoid metabolism, including those encoding phenylalanine ammonia lyase, cinnamate-4–hydroxylase and 4–coumarate CoA ligase, three genes for hydroxycinnamate biosynthesis, including those encoding hydroxycinnamoyl CoA shikimate/quinate hydroxycinnamoyltransferase and CoA:quinate hydroxycinnamoyl transferase, and 16 flavonoid biosynthetic genes, including those encoding chalcone synthase, chalcone isomerase, flavonol-*O*–methyltransferase, flavanone-3–hydroxylase, flavonoid-3′-hydroxylase, flavonoid-3′5′–hydroxylase, flavonol synthase, dihydroflavonol reductase, anthocyanidin synthase and flavonoid-3–glycosyltransferase (Figure 2b).

All genes involved in flavonoid anthocyanin biosynthesis were up-regulated (5–4088-fold and 5–1820-fold changes in B + 1w and B + 4w, respectively) with the exception of the flavonoid-3′-hydroxylase and flavonol synthase, which are involved in the flavonol-specific biosynthetic branch, and flavonol-*O*-methyltransferases, which are involved in quercetin and cyanidin production and in *O*–methylation of trichome-specific flavonol aglycones. Furthermore, genes that are not involved in anthocyanin biosynthesis, such as the hydroxycinnamate biosynthetic gene CoA:quinate hydroxycinnamoyl transferase, were not up-regulated.

We also analysed individual genes up-regulated at different developmental stages. Table 2 shows the genes up-regulated in *Del*/*Ros1* fruit at both B + 1 and B + 4 stages of fruit ripening. Gene counts were assessed as described by Rallapalli *et al.* (2014) and varied more than eightfold in purple fruit compared to WT. Twelve transcripts (nine flavonoid biosynthetic genes and three miscellaneous genes),

Table 1 GC/MS profiling of fruit tissues of MicroTom WT and the *Del/Ros1* transgenic line

	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***
Cysteine	15 ± 5	14 ± 2	0.91	14 ± 3	8 ± 2	0.59*	8 ± 1	10 ± 2	1.22
γ–aminobutyric acid	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*
Leucine	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									
2-oxoglutarate (2OG)	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***

(continued)

Table 1. (continued)

	Whole fruit			Flesh			Peel		
	WT	*Del/Ros1*	FC	WT	*Del/Ros1*	FC	WT	*Del/Ros1*	FC
Putative galacturonate	68 ± 26	58 ± 10	0.85	45 ± 9	36 ± 9	0.79	52 ± 12	59 ± 12	1.14
Putative galactonate	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

Asterisks indicate statistically significant differences between WT and the transgenic line (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$).

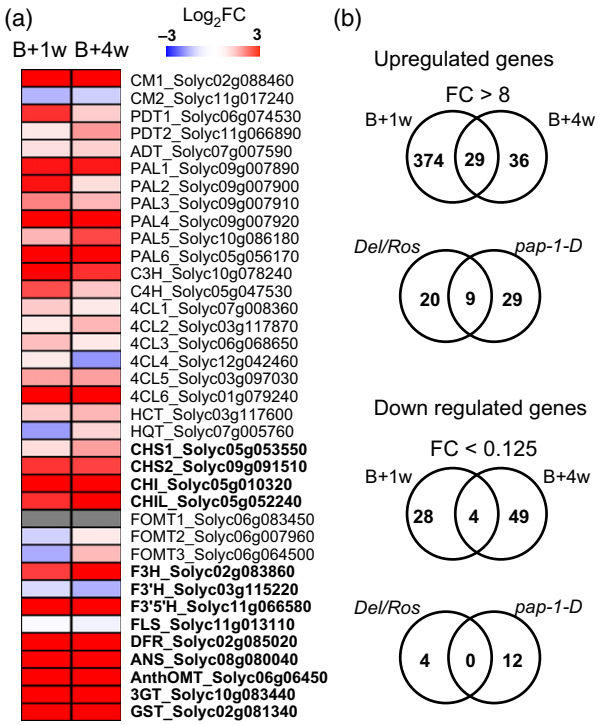


Figure 2. Gene expression profiles in *Del*/*Ros1* tomatoes.
Gene expression profiling was performed by RNA-seq for two time points: ripe stage (breaker + 1 week, B + 1w) and over-ripe stage (breaker + 4 weeks, B + 4w).
(a) The expression level of phenylpropanoid biosynthetic genes characterized or annotated previously. The intensity of log-scaled fold change (−3 to 3) is indicated using colour; red, up-regulated in *Del*/*Ros1* tomatoes; blue, down-regulated in *Del*/*Ros1* tomatoes.
(b) Venn diagrams of genes that are up-regulated (>8.0-fold) and down-regulated (<0.125-fold) at both developmental stages.

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

We compared the up-regulated genes with the genes reported to be up-regulated in the *Ant1* over-expression mutant, which showed intense purple pigmentation in seedlings (Mathews *et al.*, 2003) (Table 2). Based on the ITAG2.3 annotation, sequence of EST contigs were blasted and converted to Solyc IDs. Of the up-regulated genes in the *Del*/*Ros1* line, chalcone isomerase, a putative anthocyanin-3–*O*-glucosyltransferase, a putative anthocyanin-5–*O*-glucosyltransferase, glutathione-*S*–transferase and flower flavonoid transporter were also up-regulated in the *Ant1* over-expression experiments. We also used this dataset to look for conserved changes between the two ripening stages.

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

Characterization of anthocyanins in tomato fruit

To assess pathway structure and regulation more fully, detailed structural data for the constituent metabolites were necessary. The six most abundant anthocyanins in tomato were separated and characterized using fruit material from the purple tomato line. The crude extract was processed by preparative HPLC, and peaks corresponding to the main compounds detectable at 535 nm were collected (Figure S4 and Doc. 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 (Figure S5 and Doc. S1). The molecular ions $[M]^+$ at *m*/*z* 919 and 933 had the molecular formulae $C_{42}H_{47}O_{23}$ and $C_{43}H_{49}O_{23}$, respectively.

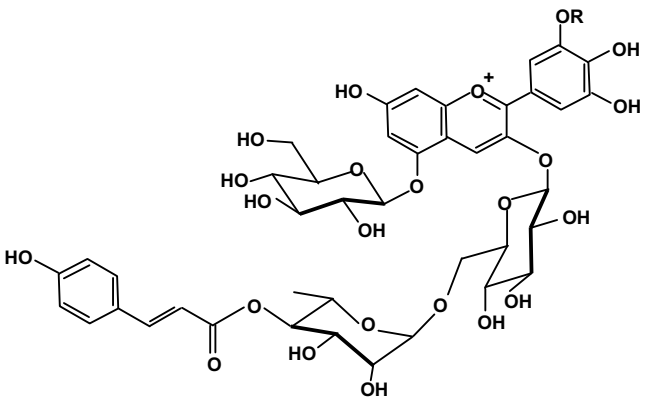
The structures of the two main anthocyanins in tomato (TA1 and TA2) were elucidated using NMR (Doc. S2). In the ^{1}H–NMR spectrum of TA1, the most significant feature was the absence of a methoxy group singlet at δ = 4.00 ppm (present in compound V), attached to the C3′ position. The NMR spectra also confirmed the presence of two glucose molecules, one rhamnose molecule and one molecule of *p*–coumaric acid in each anthocyanin. In the *p*–coumaric acid moiety, the 2″ and 3″ protons had large coupling constants (*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 moiety, the triplet signal for H4 appeared to be shifted down-field, 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

Table 2 Up-regulated genes in *Del/Ros1* tomato

ITAG2.3 ID	Arabidopsis orthologues	Gene name	Function	WT		*Del/Ros1*					
				B + 1w Count	B + 4w Count	B + 1w Count	B + 4w Count	B + 1w FC	B + 4w FC	Up-regulated	Up-regulated
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	Putative anthocyanin-3–*O*-glucosyltransferase	52	7	1960	575	38	80	*ant1*	*pap1–D*
Solyc09g059170	At5g54010	A3G2″GT	Putative anthocyanin-3–*O*-glc-2″–*O*-glucosyltransferase	12	1	1628	130	135	145		
Solyc12g098590	At4g14090	A5GT	Putative anthocyanin-5–*O*-glucosyltransferase	5	1	4978	1291	926	1441	*ant1*	*pap1–D*
Solyc12g088170	At3g26040	AAT	Putative 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	SIPAL4	Phenylalanine ammonia lyase	8	3	77	42	10	16		
Solyc05g056170	At3g53260	SIPAL6	Phenylalanine ammonia lyase	4	2	75	38	18	17		
Solyc01g079240	At2g47240	SI4CL6	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 b_5	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		

glc, glucoside.

(a)



TA1: R=H, delphinidin 3-*O*-[4-(*trans*-*p*-coumaroyl)-α-**L**-rhamnosyl-(1->6)-β-**D**-glucoside]-5-*O*-β-**D**-glucoside

TA2: R=CH_3, petunidin 3-*O*-[4-(*trans*-*p*-coumaroyl)-α-**L**-rhamnosyl-(1->6)-β-**D**-glucoside]-5-*O*-β-**D**-glucoside

(b)

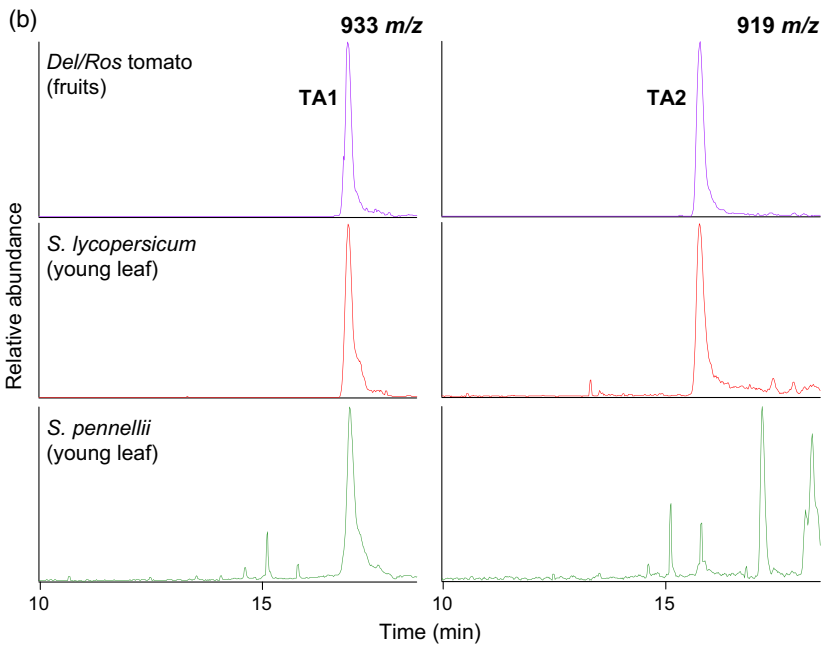


Figure 3. Characterization of major anthocyanins in tomato.
(a) Tomato anthocyanins characterized in this study.
(b) Co-elution profiles of purple tomato extracts for peak characterization of endogenous anthocyanins in leaves of domesticated tomato (*S. lycopersicum*, M82) and wild tomato (*S. pennellii*). LC/MS extracted ion chromatograms showing TA1 (petanin, *m/z* 933) and TA2 (nasunin *m/z* 919) (positive ion detection).

heteronuclear multiple bond correlation spectrum between H4 (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 values of approximately 5.5 and 5.2 ppm, respectively (glucose A and B) with a *J* value of approximately 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 of approximately 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 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 ectopic expression of snapdragon transcription factors in tomato fruits leads to the synthesis of nasunin and petanin pigments, we next analysed the anthocyanins present in tomato leaves of the common domesticated tomato line M82. As shown in Figure 3(b), young leaves of M82 contain the same anthocyanins as

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 (Koenig *et al.*, 2013) revealed that anthocyanin biosynthetic genes are more highly expressed in seedlings and mature leaves than in roots of both *S. lycopersicum* and *S. pennellii* (Figure S6) (Koenig *et al.*, 2013). Thus the expression levels are in keeping with the metabolite contents observed. Our results suggested that there was conservation of anthocyanin decoration between 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 (1999), to assess the distribution of these specific flavonoids further. We searched for all possible 4‴–*O*-phenylacylated anthocyanin-3–*O*-rutinoside-5–*O*-glucosides and for phenylacylated flavonol-3–*O*-rustinosides as summarized in Table 3. Amongst the decorated anthocyanins listed in the databases, 37 are phenylacylated delphinidin derivatives, four of which contain the -*O*–(4‴–phenylacyl-rutinoside)-5–*O*-glucoside moiety, as for 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. 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 violanin) is particularly well characterized eggplant (S. melongena; Kuroda and Wada, 1933; Sakamura *et al.*, 1963), as well as potato (Goto *et al.*, 1978), pepper (Sadilova *et al.*, 2006) and petunia (Schram *et al.*, 1983) although it is also present in Iris tingitana (Harborne, 1964; Nerdal and Andersen, 1992) and Viola tricolor (Harborne and Baxter, 1999), whereas petanin is particularly well characterized in petunia (Schram *et al.*, 1983), as has been described in potato, Solanum nigrum and Iris (Harborne and Baxter, 1999). Other anthocyanin derivatives exhibiting 3–*O*-(4‴–phenylacyl-rutinoside)-5–*O*-glucoside decorations have mostly been described in Solanaceous species (petunia species, eggplant, potato and *S. nigrum*), but also with additional species of iris, *Ipomoea indivisia* and *Silene dioica* (Table 3). These results emphasize the conservation of metabolic decoration of anthocyanins in the form of 3–*O*-(4‴–phenylacyl-rutinoside)-5–*O*-glucoside derivatives in Solanaceous species. Despite this metabolite conservation of anthocyanin-3–*O*-(4‴–phenylacyl-rutinoside)-5–*O*-glucoside derivatives, flavonol derivatives with 3–*O*-(4‴–phenylacyl)-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 RNA-seq 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 up-regulated at the transcript level in purple tomatoes, represent very strong candidates for involvement in anthocyanin decoration in tomato. Flavonoid glycosyltransferase genes most commonly belong to the UGT1 gene family. The phylogenetic analysis shown in Figure 4(a) demonstrates that the candidate anthocyanidin glycosyltransferases from tomato fall into distinct clades within the UGT1 family, with the tomato glycosyltransferase being most similar to one from petunia in each subclades. The candidate genes encoded proteins belonging to three classes, namely a flavonoid-3–*O*-glycosyltransferase (SlFd3GlcT; Solyc10g083440), an anthocyanin-5–*O*-glycosyltransferase (SlA5GlcT; Solyc12g098590) and a flavonoid-3–*O*-glycoside-*O*–glycosyltransferase (SlA3Glc6″RhaT; Solyc09g059170). This sequence similarity to other genes from both within the Solanaceous family and 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 as a member of the BAHD family, but not in the subclades of known functions (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 the eFP browser (http://bar.utoronto.ca/efp_tomato/cgi-bin/efpWeb.cgi) (Figure S7).

As a further step towards understanding the anthocyanin metabolic pathway, we searched the PLAZA database (http://bioinformatics.psb.ugent.be/plaza) for orthologues of tomato genes known, or suggested from our study, to encode each step of the anthocyanin biosynthetic pathway. The results of this search are presented in Figure 4(c), which indicates the number of genes found for each enzyme and the number of species in which they were found. For the majority of enzymatic steps, genes were found in more than 20 species, and, for some, enzymatic reactions are obviously catered for by a large number of

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

Flavonoid*		Formula	Species	Plant name
Anthocyanins				
Delphinidin derivatives (37)				
C00006893	Delphinidin-3–(4‴-*trans*-*p*–coumaroyl-Rut)-5–Glc Synonyms: violanin, nasunin	$C_{42}H_{47}O_{23}$	Iridaceae Solanaceae Solanaceae Solanaceae Solanaceae Violaceae	*Iris tingitana* *Petunia* x *hybrida* *Solanum melongena* *Solanum tuberosum* *Capsicum annuum* *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 Synonym: petanin	$C_{43}H_{49}O_{23}$	Iridaceae Solanaceae Solanaceae Solanaceae	*Iris* spp. *Petunia* x *hybrida* *Solanum nigrum* *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 Solanaceae	*Petunia guarapuavensis* *Petunia hybrida* cultivar
C00011075	Malvidin-3–(4‴-(6⁗⁗–caffeoyl-Glc)-caffeoyl)-Rut-5–Glc	$C_{59}H_{67}O_{32}$	Solanaceae Solanaceae	*Petunia guarapuavensis* *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 Synonym: peonanin	$C_{43}H_{49}O_{22}$	Solanaceae Solanaceae Solanaceae Solanaceae	*Petunia* x *hybrida* *Solanum nigrum* *Solanum phureja* *Solanum tuberosum*
C00006874	Peonidin-3–(4‴-caffeoyl)-Rut-5–Glc	$C_{43}H_{49}O_{23}$	Convolvulaceae Solanaceae Solanaceae	*Ipomoea indivisa* *Petunia* x *hybrida* *Solanum tuberosum*
Pelargonidin derivatives (41)				
C00006789	Pelargonidin-3–(4‴-*p*–coumaryl)-Rut-5–Glc Synonym: pelanin	$C_{42}H_{47}O_{21}$	Caryophyllaceae Solanaceae	*Silene dioica* *S. andigena* × *S. tuberosum*
C00014853	Pelargonidin-3–(4‴-feruloyl)-Rut-5–Glc	$C_{43}H_{49}O_{22}$	Solanaceae	*S. andigena* ×*S. 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				

*The total number of phenylacylated flavonoids is shown in parentheses.
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 (1999) for information on these flavonoids. References are supplied in Doc. S3.

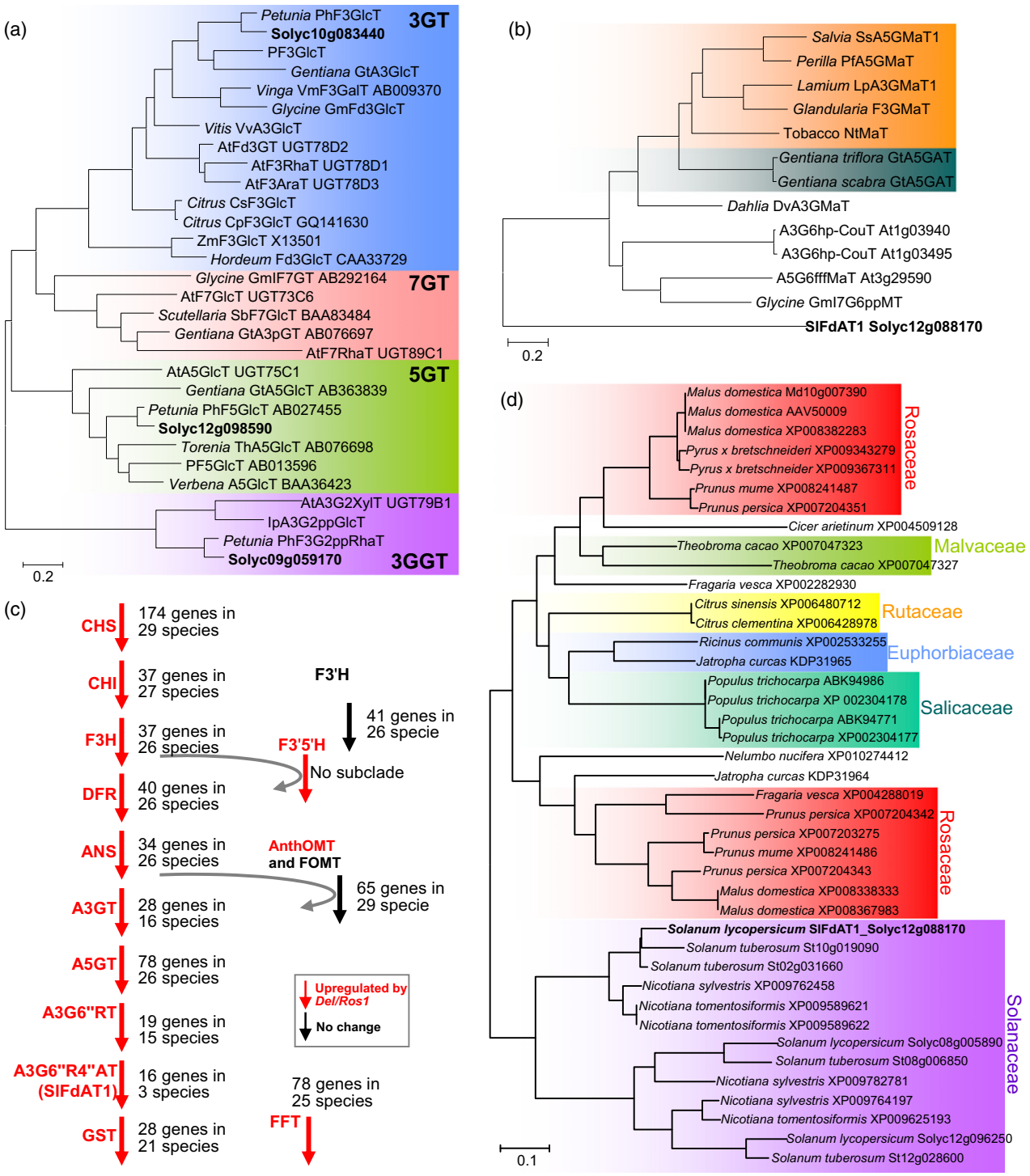


Figure 4. Phylogenetic analysis of candidate UGT and BAHD genes involved in flavonoid biosynthesis.
(a, b) Molecular phylogenetic trees of the amino acid sequences of the (a) flavonoid glycosyltransferases and (b) flavonoid acyltransferases from tomato. The amino acid sequences were aligned using MEGA5.1 (http://www.megasoftware.net/).
(c) A total of 31 plant species were used for orthologue gene cluster analysis using PLAZA 3.0 Dicots (http://bioinformatics.psb.ugent.be/plaza//). Red arrows indicate up-regulation in *Del*/*Ros1* tomatoes.
(d) Molecular phylogenetic tree of amino acid sequences corresponding to ORFs encoding orthologues to SlFdAT1 (Solyc12g088170). The amino acid sequences were aligned using MEGA5.1 (http://www.megasoftware.net/).

isoforms. There were four exceptions with respect to number of species: flavonoid-3′5′-hydroxylase (F3′5′H), anthocyanin-3–*O*-glucosyltransferase (A3GlcT), anthocyanin-3–*O*-glucoside-6″–*O*-rhamnosyltransferase (A3Glc6″RhaT) and flavonoid-3–*O*-rutinoside-4‴–*O*-phenylacyltransferase (SlFdAT1) (Figure 4c). As SlFdAT1 shows very poor conservation and is limited to three species, we searched gene sequences submitted to Genbank for orthologues of SlFdAT (Figure 4d). Targeted phylogenetic analysis showed a clear separation between a BAHD subclade that included SlFdAT1 (http://www.ncbi.nlm.nih.gov/) and ATs from the Solanaceae, such as potato andtobacco, but did not include ATs from other plant families (Figure 4d).

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 SOL database (http://solgenomics.net/) 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 *Escherichia coli*, and total protein was measured in crude extracts. 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, 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 acyl acceptors (Figure 5a,b) and flavonol acyl acceptors (Figure 5c,d) and all phenylacyl

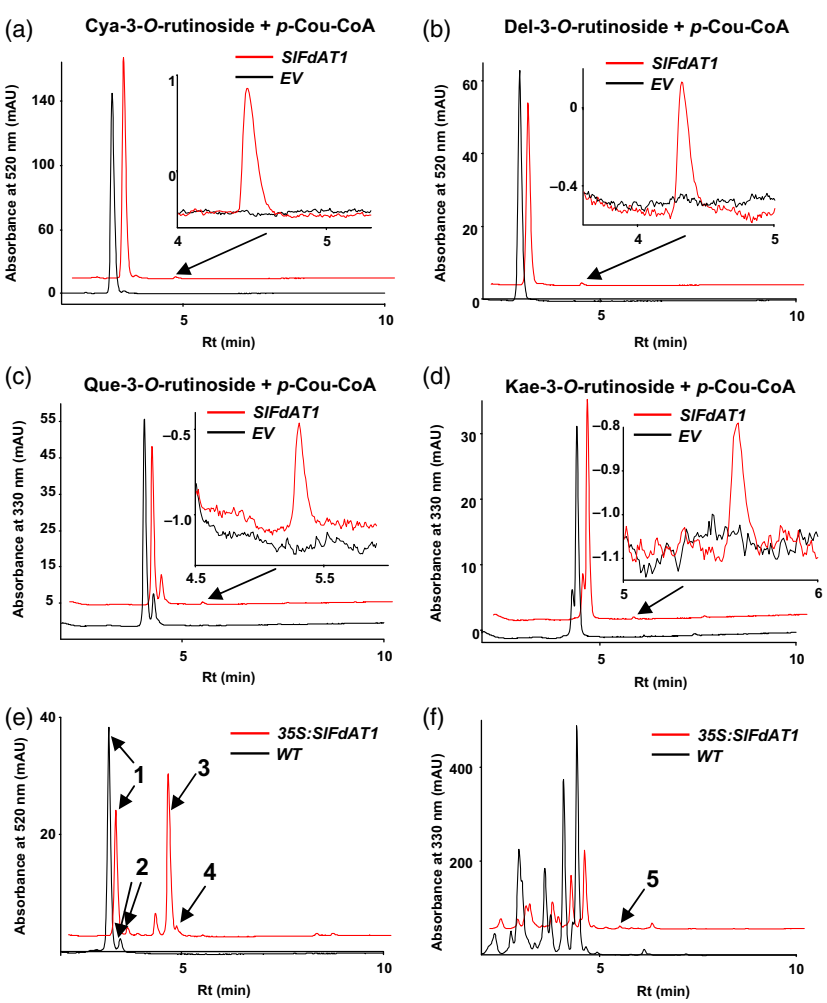


Figure 5. Functional characterization of SlFdAT1.
(a–d) HPLC chromatograms for assays of the SlFdAT acyltransferase from *S. lycopersicum*.
(a) Activity of recombinant protein expressed in *E. coli* with cyanidin-3–*O*-rutinoside as acyl acceptor and *p*–coumaroyl CoA as acyl donor, (b) delphinidin-3–*O*-rutinoside 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, and (d) kaempferol-3–*O*-rutinoside as acyl acceptor and *p*–coumaroyl CoA as acyl donor.
(e, f) HPLC chromatograms of flower extracts of WT and the *35S:SlFdAT1* tobacco line showing common peaks (peak 1, 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.

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

	nmol $sec^{-1}mg^{-1}$	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 performed using *p*–coumaroyl CoA as the acyl donor.
[c]The reactions were performed using cyanidin 3–rutinoside as the acyl acceptor.

donors, but no activity with malonyl CoA as the acyl donor (Table 4).

Additionally, *SlFdAT1* was expressed in tobacco under the control of the 35S promoter to examine SlFdAT1 function *in vivo* (Figure 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 line over-expressing *SlFdAT1* showed production of significant levels of phenylacylated cyanidin-3–*O*-rutinosides and trace levels of phenylacylated quercetin-3–*O*-rutinoside (Figure 5f,g and Figure S8). Given the results of the assays of recombinant Sl*Fd*AT1 protein *in vitro*, these data confirm that SlFdAT1 acts as a flavonoid-3–*O*-rutinoside-4‴–*O*-phenylacyltransferase *in vivo*, and its activity accounts 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

Over-expression 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 that they regulate (Sharma and Dixon, 2005; Tohge *et al.*, 2005; Peel *et al.*, 2009; Zhao *et al.*, 2010; Liu *et al.*, 2014). Here we performed an in-depth analysis of *Del*/*Ros1* purple tomatoes that accumulate anthocyanins at concentrations comparable to those found in blackberries (*Rubus* spp.) and blueberries (*Vaccinium* spp.), enhancing the antioxidant potential of the fruit 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 obtain a fuller picture of the compositional changes resulting from 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 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, 2014). These changes included depletion of phenylalanine, 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 mostly focused on anthocyanins, several 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 phenylalanine, which was observed in peel, pericarp and whole-fruit samples, as were decreased levels of alanine, threitol and galactose. Phenylalanine is the direct precursor of the phenylpropanoid pathway, and its decreased levels are probably the direct result of increased demand for anthocyanin biosynthesis. Interestingly, analysis of the RNA-seq data revealed that this decrease occurred despite elevated expression of genes involved in phenylalanine biosynthesis, suggesting that the increased demand for phenylalanine 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. As RNA-seq analysis showed up-regulation of genes involved in not only flavonoid biosynthesis but also phenylpropanoid, phenylalanine and shikimate biosynthesis, over-accumulation of quinate and CGAs are probably the result of enhanced demand for flavonoids in the *Del*/*Ros1* line.

Structural elucidation of tomato anthocyanins has previously proven very difficult, as they are not usually available in large amounts (Jones *et al.*, 2003), and prior to this study, accumulation as a result of genetic manipulation had rarely been detected in tomato flesh (Bovy *et al.*, 2002). A low abundance of anthocyanins in tissues is observed in tomatoes and in most non-domesticated tomato accessions such as *S. pennellii*. Using the high-yielding purple tomatoes, six 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.

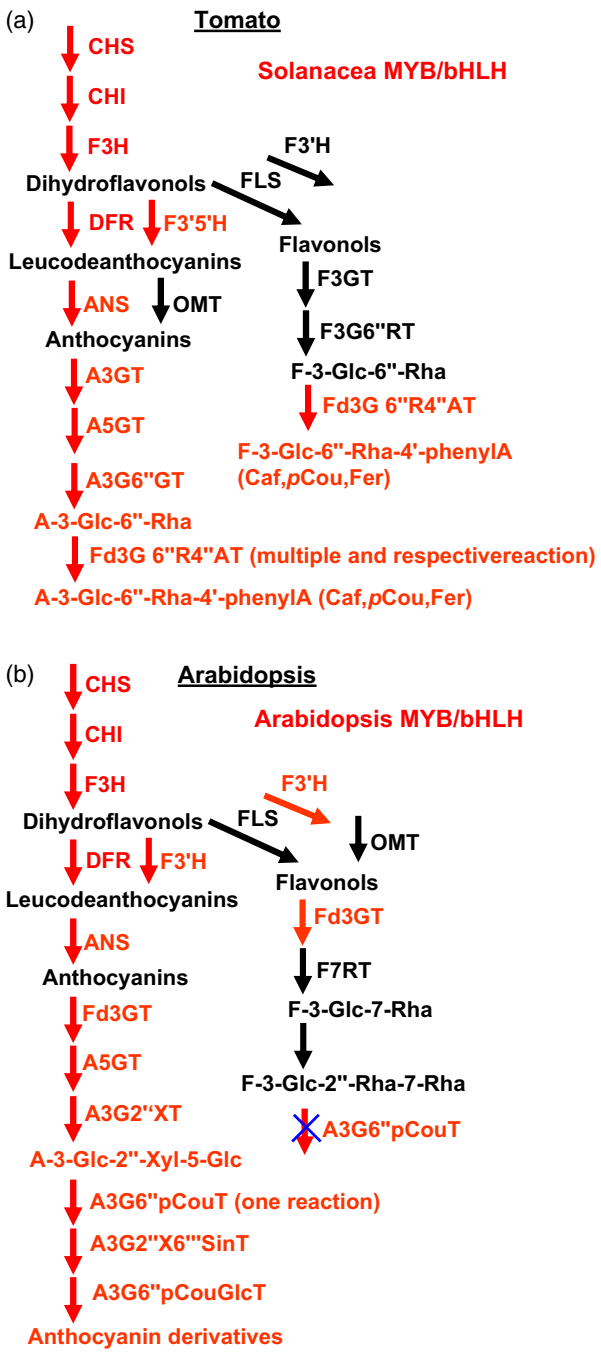


Figure 6. Summary of differences in regulation of anthocyanin biosynthesis between tomato and Arabidopsis.
Red arrows indicate up-regulation in (a) *Del*/*Ros1* tomato or (b) *pap1*–*D* Arabidopsis.

Genome-scale profiling of anthocyanin biosynthesis has already been reported for Arabidopsis using the mutant *pap1*–*D*, which over-expresses 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 that enable up-regulation of genes involved in the first steps of the anthocyanin biosynthetic pathway (early biosynthetic genes), which are common to most angiosperms, and some late biosynthetic genes, 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). However, in such mutants, expression of early biosynthetic genes may be complemented by the activity of other MYB regulators, such as members of the MYB family (subgroup 7). This means that early biosynthetic genes may be mis-scored as not regulated by the MBW complex, of which DEL and ROS1 are part (Zhang *et al.*, 2014).

As anthocyanin acyl acceptors are not exactly the same in Arabidopsis (anthocyanin-3–*O*-glucoside) and Solanaceous species (anthocyanin-3–*O*-rutinoside), differences may be observed in late biosynthetic genes such as glycosyltransferases and acyltransferases for the two species. The dual functionality of SlFdAT1 in catalysing 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 Solanaceous 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 (Pichersky and Lewinsohn, 2011). 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 SlFdAT1, are very uncommon in the plant kingdom. SlFdAT1 may 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, i.e. 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 (GLABRA3 (GL3), ENHANCER OF GLABROUS 3 (EGL3) and 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 Solanaceae family. In Solanaceous species such as tobacco, which lacks 5–*O*-glycosylation of its anthocyanins, the gene encoding the A5GlcT has probably been lost. In contrast, the gene encoding SlFdAT1 appears to have arisen after divergence of the Solanaceae, and must have recruited the MBW complex 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 (Kochevenko *et al.*, 2012). Whole fruits and two tissues (flesh and peel) were harvested at two stages (breaker + 1 week and breaker + 4 weeks) (Zhang *et al.*, 2013). 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 performed by GC/MS as described by Lisec *et al.* (2006), and by LC/MS by extracting metabolites as described by Tohge and Fernie (2010) and analysing these extracts as described by Rohrmann *et al.* (2011) and Schwahn *et al.* (2014). Information concerning metabolite identification is provided based on reporting suggestions for large-scale metabolite datasets (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 the P7 sequence Illumina (http://www.illumina.com/). Double-strand cDNA was synthesized as described previously (Okayama and Berg, 1982). Purified cDNA was subjected to Covaris shearing (http://covarisinc.com/) (intensity 5, duty cycle 20%, cycles/burst 200, duration 90 sec). End repairing and A–tailing of sheared cDNA was performed as described by Illumina. Y–shaped adapters were ligated to A–tailed DNA and subjected to size selection on agarose gels. The libraries were sequenced on an Illumina Genome Analyzer IIx. The Illumina sequence library was quality filtered using FASTX-Toolkit 0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/index.html) with parameters -q20 and -p50. The sub-library was artefact-filtered using the FASTX Toolkit. Quality-filtered libraries were aligned ITAG2.3 *S. lycopersicum* cDNA sequences (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 the tomato genome: (ftp://ftp.solgenomics.net/tomato_genome/annotation/ITAG2.3_release/ITAG2.3_genomic.fasta) using BOWTIE version 0.12.8. The RNA-seq data has been deposited in the National Center for Biotechnology Information Gene Expression Omnibus (GEO) under accession number GSE61014.

Isolation and purification of the anthocyanins

For isolation of the major anthocyanins in *Del*/*Ros1* tomato fruit, extracts were prepared from 5 g powdered tomato fruit using 50 ml of 50% aqueous MeOH and 25 ml of 100% MeOH. The extract was clarified by paper filtration and then concentrated in a rotary evaporator to a final volume of 20 ml and purified by preparative HPLC. HPLC conditions are described in Doc. S1, S2, S3.

Identification of acyl moieties and sugar moieties by hydrolysis

For identification of acyl moieties, 20 μl of pure anthocyanin fractions were subjected to alkaline hydrolysis with 250 μl of 10% KOH for 30 min at room temperature. The hydrolysate was acidified to pH 1.0 with 250 μl of 2 N HCl, and the saponification products were extracted using three volumes ethyl acetate. For identification of sugar moieties, 20 μl of pure anthocyanin fractions were subjected to acid hydrolysis with 120 μl of 2 N HCl for 30 min at 95°C in a sealed vial. After cooling of the hydrolysate in an ice bath, hydrolyzedacyl moieties were extracted using 1 ml of 1–pentanol. The fractions obtained were dried and resuspended for LC/MS. The LC/MS method for profiling of hydrolysates is described in Doc. S1.

NMR to identify anthocyanins

^{1}H- and ^{13}C-NMR spectra were acquired using a Bruker (https://en.wikipedia.org/wiki/Bruker) DMX 500 NMR spectrometer operating at 500.13 MHz for ^{1}H–NMR and 125.77 MHz for ^{13}C–NMR. CD_3OD/CF_3COOD (10:1) solutions containing approximately 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 and 49.00 ppm for residual CHD_2OD and CD_3OD signals, respectively. ^{1}H- and ^{13}C-NMR spectral data are shown in Doc. S2.

In vitro assay of recombinant SlFdAT1 protein

Enzymatic assay of SlFdAT1 (Solyc12g088170) using recombinant protein was performed using a modification of the method described by Luo *et al.* (2007). Full-length *SlFdAT1* (Solyc12g088170) cDNA was amplified from *Del*/*Ros1* tomato fruit cDNA using primers *SlFdAT1*-attB1 (5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTGGATGAGCCAAATTACAACACAAAA-3′) and *SlFdAT1*-attB2 (5′-GGGGACCACTTTGTACAAGAAAGCTGGGTCCTACTACTTTGGCACATAACTA-3′).

PCR products were recombined with pDONR207 vectors Luo *et al.*, 2007 to create pENTR207-*SlFdAT1*. After sequence verification, the entry clone was introduced into the N–terminal S–TAG fusion vector pJAM1784 to create pJAM1784-*SlFdAT1* (Luo *et al.*, 2007) . Recombinant protein expression and purification were performed as described previously (Luo *et al.*, 2007). The S–TAG protein concentration was determined using an S–TAG rapid assay kit (Novagen; http://www.merckmillipore.com/) according to the manufacturer's instructions. Enzyme assays were performed 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 (http://www.plantmetachem.com/). Quercetin-3–*O*-rutinoside (rutin) and kaempferol-3–*O*-rutinoside were purchased from Sigma (https://www.sigmaaldrich.com). Delphinidin-3–*O*-rutinoside was purchased from EXTRASYNTHESE (http://www.extrasynthese.com/). The LC/MS method used for the detection of enzymatic products is described in Doc. S3.

Ectopic expression of SlFdAT1 in tobacco

The *SlFdAT1* (Solyc12g088170) full-length cDNA was amplified using primers SlFdAT1-attB1 and SlFdAT1-attB2, and cloned into a pBin19–GW binary vector (Butelli *et al.*, 2012) by 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 the transgenic line were harvested and used for LC/MS analysis.

ACKNOWLEDGEMENTS

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

SUPPORTING INFORMATION

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

Figure S1. LC/MS total ion chromatogram of WT and *Del*/*Ros1* tomatoes in flesh and peel.

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

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

Figure S4. Semi-preparative HPLC profile at 535 nm for the main anthocyanins from the *Del*/*Ros1* transgenic tomato.

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

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

Figure S7. eFP gene expression analysis.

Figure S8. MS/MS analysis of major anthocyanins found in *35S: SlFdAT1* transgenic tobacco flowers.

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

Table S2. RNA-seq data for WT and the *Del*/*Ros1* transgenic tomato line.

Doc S1. HPLC methods.

Doc S2. ^{1}H- and ^{13}C-NMR spectral data for compounds III and V.

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

REFERENCES

Afendi, F.M., Okada, T., Yamazaki, M. *et al.* (2012) KNApSAcK family databases: integrated metabolite–plant species databases for multifaceted plant research. *Plant Cell Physiol.* **53**, e1.

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.

Aviram, M. and Fuhrman, B. (2002) Wine flavonoids protect against LDL oxidation and atherosclerosis. *Ann. N. Y. Acad. Sci.* **957**, 146–161.

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.

Borevitz, J.O., Xia, Y.J., Blount, J., Dixon, R.A. and 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. *et al.* (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, Z.T., Libault, M., Findley, S., Sugawara, M., Sadowsky, M.J., Sumner, L.W. and Stacey, G. (2010) Soybean metabolites regulated in root hairs in response to the symbiotic bacterium *Bradyrhizobium japonicum*. *Plant Physiol.* **153**, 1808–1822.

Butelli, E., Titta, L., Giorgio, M. *et al.* (2008) Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. *Nat. Biotechnol.* **26**, 1301–1308.

Butelli, E., Licciardello, C., Zhang, Y., Liu, J., Mackay, S., Bailey, P., Reforgiato-Recupero, G. and Martin, C. (2012) Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. *Plant Cell*, **24**, 1242–1255.

Carrari, F., Baxter, C., Usadel, B. *et al.* (2006) Integrated analysis of metabolite and transcript levels reveals the metabolic shifts that underlie tomato fruit development and highlight regulatory aspects of metabolic network behavior. *Plant Physiol.* **142**, 1380–1396.

Commenges, D., Scotet, V., Renaud, S., Jacqmin-Gadda, H., Barberger-Gateau, P. and Dartigues, J.F. (2000) Intake of flavonoids and risk of dementia. *Eur. J. Epidemiol.* **16**, 357–363.

Dal Cin, V., Tieman, D.M., Tohge, T. *et al.* (2011) Identification of genes in the phenylalanine metabolic pathway by ectopic expression of a MYB transcription factor in tomato fruit. *Plant Cell*, **23**, 2738–2753.

D'Auria, J.C. (2006) Acyltransferases in plants: a good time to be BAHD. *Curr. Opin. Plant Biol.* **9**, 331–340.

Enfissi, E.M., Barneche, F., Ahmed, I. *et al.* (2010) Integrative transcript and metabolite analysis of nutritionally enhanced DE-ETIOLATED1 downregulated tomato fruit. *Plant Cell*, **22**, 1190–1215.

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.

Fitzpatrick, T.B., Basset, G.J.C., Borel, P. *et al.* (2012) Vitamin deficiencies in humans: can plant science help? *Plant Cell*, **24**, 395–414.

Grotewold, E., Sainz, M.B., Tagliani, L., Hernandez, J.M., Bowen, B. and Chandler, V.L. (2000) Identification of the residues in the Myb domain of maize C1 that specify the interaction with the bHLH cofactor R. *Proc. Natl Acad. Sci. USA*, **97**, 13579–13584.

Gutteridge, J.M.C. (1994) Biological origin of free radicals, and mechanisms of antioxidant protection. *Chem. Biol. Interact.* **91**, 133–140.

Halliwell, B. (2012) Free radicals and antioxidants: updating a personal view. *Nutr. Rev.* **70**, 257–265.

Harborne, J.B. (1964) Plant polyphenols. 11. the structure of acylated anthocyanins. *Phytochemistry*, **3**, 151–160.

Harborne, J.B. and Baxter, H. (1999) *Handbook of Natural Flavonoids*. Chichester: Wiley, pp. 1–2.

Hirai, M.Y., Sugiyama, K., Sawada, Y. *et al.* (2007) Omics-based identification of Arabidopsis Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *Proc. Natl Acad. Sci. USA*, **104**, 6478–6483.

Hollman, P.C.H., Hertog, M.G.L. and Katan, M.B. (1996) Analysis and health effects of flavonoids. *Food Chem.* **57**, 43–46.

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

Jones, C.M., Mes, P. and Myers, J.R. (2003) Characterization and inheritance of the anthocyanin fruit (Aft) tomato. *J. Hered.* **94**, 449–456.

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

Kochevenko, A., Klee, H.J., Fernie, A.R. and Araujo, W.L. (2012) Molecular identification of a further branched-chain aminotransferase 7 (BCAT7) in tomato plants. *J. Plant Physiol.* **169**, 437–443.

Koenig, D., Jimenez-Gomez, J.M., Kimura, S. *et al.* (2013) Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *Proc. Natl Acad. Sci. USA*, **110**, E2655–E2662.

Kong, Q., Pattanaik, S., Feller, A., Werkman, J.R., Chai, C., Wang, Y., Grotewold, E. and Yuan, L. (2012) Regulatory switch enforced by basic helix-loop-helix and ACT-domain mediated dimerizations of the maize transcription factor R. *Proc. Natl Acad. Sci. USA*, **109**, E2091–E2097.

Kuroda, C. and Wada, M. (1933) The colouring matter of eggplant (Nasu). Part I. *Proc. Imp. Acad.* **9**, 51–52.

Kusano, M., Tohge, T., Fukushima, A. *et al.* (2011) Metabolomics reveals comprehensive reprogramming involving two independent metabolic responses of Arabidopsis to UV-B light. *Plant J.* **67**, 354–369.

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

Le Gall, G., DuPont, M.S., Mellon, F.A., Davis, A.L., Collins, G.J., Verhoeyen, M.E. and Colquhoun, I.J. (2003) Characterization and content of flavonoid glycosides in genetically modified tomato (*Lycopersicon esculentum*) fruits. *J. Agric. Food Chem.* **51**, 2438–2446.

Le Marchand, L. (2002) Cancer preventive effects of flavonoids – a review. *Biomed. Pharmacother.* **56**, 296–301.

Lepelley, M., McCarthy, J.G., Petiard, V., Cheminade, G., Tanksley, S.D. and Lin, C. (2012) *Polynucleotides encoding lignin biosynthetic pathway enzymes in coffee.* Patent number PCT/US06/040223.

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

Liu, Z., Shi, M.-Z. and 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., Nishiyama, Y., Fuell, C. *et al.* (2007) Convergent evolution in the BAHD family of acyl transferases: identification and characterization of anthocyanin acyl transferases from *Arabidopsis thaliana*. *Plant J.* **50**, 678–695.

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

Luo, J., Fuell, C., Parr, A., Hill, L., Bailey, P., Elliott, K., Fairhurst, S.A., Martin, C. and Michael, A.J. (2009) A novel polyamine acyltransferase responsible for the accumulation of spermidine conjugates in Arabidopsis seed. *Plant Cell*, **21**, 318–333.

Martin, C. (2013) The interface between plant metabolic engineering and human health. *Curr. Opin. Biotechnol.* **24**, 344–353.

Martin, C., Zhang, Y., Tonelli, C. and Petroni, K. (2013) Plants, diet, and health. *Annu. Rev. Plant Biol.* **64**, 19–46.

Mathews, H., Clendennen, S.K., Caldwell, C.G. *et al.* (2003) Activation tagging in tomato identifies a transcriptional regulator of anthocyanin biosynthesis, modification, and transport. *Plant Cell*, **15**, 1689–1703.

Mojzisova, G. and Kuchta, M. (2001) Dietary flavonoids and risk of coronary heart disease. *Physiol. Res.* **50**, 529–535.

Mounet, F., Moing, A., Garcia, V. *et al.* (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 Physiol.* **149**, 1505–1528.

Muir, S.R., Collins, G.J., Robinson, S., Hughes, S., Bovy, A., De Vos, C.H.R., 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.

Nerdal, W. and Andersen, O.M. (1992) Intermolecular aromatic acid association of an anthocyanin (petanin) evidenced by 2-dimensional nuclear Overhauser enhancement nuclear magnetic resonance experiments and distance geometry calculations. *Phytochem. Anal.* **3**, 182–189.

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

Peel, G.J., Pang, Y., Modolo, L.V. and Dixon, R.A. (2009) The LAP1 MYB transcription factor orchestrates anthocyanidin biosynthesis and glycosylation in *Medicago*. *Plant J.* **59**, 136–149.

Pichersky, E. and Lewinsohn, E. (2011) Convergent evolution in plant specialized metabolism. In *Annual Review of Plant Biology* (Merchant, S.S., Briggs, W.R. and Ort, D., eds). Vol 62., Palo Alto: Annual Reviews, pp. 549–566.

Purdy, S.J., Bussell, J.D., Nunn, C.P. and Smith, S.M. (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, J.F., Leppen, H.T.C., Mol, J.N.M. and Koes, R.E. (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, E.M., Robert-Seilaniantz, A., Segonzac, C., Etherington, G.J., Sohn, K.H., MacLean, D. and Jones, J.D.G. (2014) EXPRSS: an Illumina based high-throughput expression-profiling method to reveal transcriptional dynamics. *BMC Genomics*, **15**, 341.

Ramsay, N.A. and Glover, B.J. (2005) MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. *Trends Plant Sci.* **10**, 63–70.

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.

Sakamura, S., Watanabe, S. and Obata, Y. (1963) The structure of the major anthocyanin in eggplant. *Agric. Biol. Chem.* **27**, 663–665.

Sadilova, E., Stintzing, F.C. and Carle, R. (2006) Anthocyanins, colour and antioxidant properties of eggplant (*Solanum melongena* L.) and violet pepper (*Capsicum annuum* L.) peel extracts. *Z. Naturforsch. C.* **61**, 527–535.

Schmidt, R., Mieulet, D., Hubberten, H.-M. *et al.* (2013) SALT-RESPONSIVE ERF1 regulates reactive oxygen species-dependent signaling during the initial response to salt stress in rice. *Plant Cell*, **25**, 2115–2131.

Schram, A.W., Jonsson, L.M.V. and Devlaming, P. (1983) Identification of anthocyanins and intermediates of anthocyanin biosynthesis from petunia-hybrida using high-performance liquid-chromatography. *Z. Naturforsch. C.* **38**, 342–345.

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.

Schwinn, K., Venail, J., Shang, Y.J., Mackay, S., Alm, V., Butelli, E., Oyama, R., Bailey, P., Davies, K. and 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, S.B. and Dixon, R.A. (2005) Metabolic engineering of proanthocyanidins by ectopic expression of transcription factors in *Arabidopsis thaliana*. *Plant J.* **44**, 62–75.

Slimestad, R. and Verheul, M. (2009) Review of flavonoids and other phenolics from fruits of different tomato (*Lycopersicon esculentum* Mill.) cultivars. *J. Sci. Food Agr.* **89**, 1255–1270.

Sonderby, I.E., Hansen, B.G., Bjarnholt, N., Ticconi, C., Halkier, B.A. and Kliebenstein, D.J. (2007) A systems biology approach identifies a R2R3 MYB gene subfamily with distinct and overlapping functions in regulation of aliphatic glucosinolates. *PLoS One*, **2**, e1322.

Tamagnone, L., Merida, A., Stacey, N., Plaskitt, K., Parr, A., Chang, C.F., Lynn, D., Dow, J.M., Roberts, K. and 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. 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., Watanabe, M., Hoefgen, R. and Fernie, A.R. (2013a) The evolution of phenylpropanoid metabolism in the green lineage. *Crit. Rev. Biochem. Mol. Biol.* **48**, 123–152.

Tohge, T., Watanabe, M., Hoefgen, R. and Fernie, A.R. (2013b) Shikimate and phenylalanine biosynthesis in the green lineage. *Front. Plant Sci.* **4**, 1–13.

Tohge, T., Alseekh, S. and Fernie, A. (2014) On the regulation and function of secondary metabolism during fruit development and ripening. *J. Exp. Bot.* **65**, 4599–4611.

Urbanczyk-Wochniak, E., Usadel, B., Thimm, O. *et al.* (2006) Conversion of MapMan to allow the analysis of transcript data from Solanaceous species: effects of genetic and environmental alterations in energy metabolism in the leaf. *Plant Mol. Biol.* **60**, 773–792.

Vanholme, R., Storme, V., Vanholme, B., Sundin, L., Christensen, J.H., Goeminne, G., Halpin, C., Rohde, A., Morreel, K. and Boerjan, W. (2012) A systems biology view of responses to lignin biosynthesis perturbations in Arabidopsis. *Plant Cell*, **24**, 3506–3529.

Willits, M.G., Kramer, C.M., Prata, R.T.N., De Luca, V., Potter, B.G., Steffens, J.C. and Graser, G. (2005) Utilization of the genetic resources of wild species to create a nontransgenic high flavonoid tomato. *J. Agric. Food Chem.* **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, A.D., Garapati, P. *et al.* (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. *et al.* (2012) Two glycosyltransferases involved in anthocyanin modification delineated by transcriptome independent component analysis in *Arabidopsis thaliana*. *Plant J.* **69**, 154–167.

Zhang, Y., Butelli, E., De Stefano, R. *et al.* (2013) Anthocyanins double the shelf life of tomatoes by delaying overripening and reducing susceptibility to gray mold. *Curr. Biol.* **23**, 1094–1100.

Zhang, Y., Butelli, E. and Martin, C. (2014) Engineering anthocyanin biosynthesis in plants. *Curr. Opin. Plant Biol.* **19**, 81–90.

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