

1 **LARGE-SCALE BIOLOGY ARTICLE**

2

3 **Candidate Gene Networks for Acylsugar Metabolism and Plant Defense in**  
4 **Wild Tomato *Solanum pennellii***

5

6 **Sabyasachi Mandal, Wangming Ji, Thomas D. McKnight**

7

8 Department of Biology, Texas A&M University, College Station, TX 77843, USA

9 ORCID IDs: 0000-0001-6093-3468 (S.M.); 0000-0002-9863-2035 (T.D.M.)

10 Corresponding Author: [mcknight@bio.tamu.edu](mailto:mcknight@bio.tamu.edu), +1-979-845-3896

11

12 **Short title:** Acylsugar and plant defense gene networks

13

14 The author responsible for distribution of materials integral to the findings presented in this  
15 article in accordance with the policy described in the Instructions for Authors is: Thomas D.  
16 McKnight ([mcknight@bio.tamu.edu](mailto:mcknight@bio.tamu.edu)).

17

18

19

20

21

22

1 **ABSTRACT**

2 Many plants in the Solanaceae family secrete acylsugars, which are branched-chain and straight-  
3 chain fatty acids esterified to glucose or sucrose. These compounds have important roles in plant  
4 defense and potential commercial applications. However, several acylsugar metabolic genes  
5 remain unidentified, and little is known about regulation of this pathway. We used comparative  
6 transcriptomic analysis between low- and high-acylsugar-producing accessions of *Solanum*  
7 *pennellii* and found that expression levels of most acylsugar metabolic genes, including known  
8 acylsucrose biosynthetic genes and novel candidate genes (putatively encoding a ketoacyl-ACP  
9 synthase IV/II-like enzyme, peroxisomal acyl-activating enzymes, ABC transporters, and central  
10 carbon metabolic enzymes), were positively correlated with acylsugar accumulation, except two  
11 acylglucose biosynthetic genes. Genes putatively encoding oxylipin metabolic proteins,  
12 subtilisin-like proteases, and other antimicrobial defense proteins were upregulated in low-  
13 acylsugar-producing accessions, possibly to compensate for diminished defense activities of  
14 acylsugars. Gene co-expression network analysis clustered most differentially expressed genes  
15 into two separate modules and identified genetic networks associated with acylsugar production  
16 and plant defense. Transcriptome analysis after inhibition of biosynthesis of branched-chain  
17 amino acids (precursors to branched-chain fatty acids) further supported the coordinated  
18 regulation of most acylsugar candidate genes and identified three putative AP2-family  
19 transcription factor genes that form a strong co-expression network with many acylsugar  
20 metabolic genes.

21

22

23

24

25

26

27

1

## 2 INTRODUCTION

3 Plants synthesize a wide variety of specialized metabolites that provide selective advantage in  
4 specific environmental conditions. Different classes of specialized metabolites are found in  
5 specific taxonomic groups and many of them are valuable phytochemicals. Acylsugars are  
6 nonvolatile and viscous specialized metabolites secreted through trichomes of solanaceous plants  
7 (Fobes et al., 1985; Kroumova et al., 2016; Moghe et al., 2017), and their roles in plant defense  
8 have been studied extensively. For example, acylsugars from *Solanum pennellii* act as feeding  
9 and oviposition deterrents for insect pests and exert toxic effects on different herbivores  
10 (Hawthorne et al., 1992; Juvik et al., 1994; Liedl et al., 1995). Acylsugars from other genera also  
11 have important roles in providing protection against herbivores and plant pathogens (Chortyk et  
12 al., 1997; Hare, 2005; Luu et al., 2017). Additionally, amphipathic acylsugar molecules are  
13 believed to reduce the surface tension of adsorbed dew, thereby providing an additional source of  
14 water for the plants (Fobes et al., 1985). Acylsugars also have potential applications as pesticides  
15 (Puterka et al., 2003), food additives (Hill and Rhode, 1999), cosmetic and personal care  
16 products (Hill and Rhode, 1999), antibiotics (Chortyk et al., 1993), and anti-inflammatory  
17 compounds (Perez-Castorena et al., 2010). Due to these important biological roles and  
18 commercial applications of acylsugars, a detailed understanding of genetic network involved in  
19 acylsugar metabolism is required for successful crop breeding programs and metabolic  
20 engineering of acylsugar production.

21 Acylsugars typically consist of aliphatic, branched- and/or straight-chain fatty acids  
22 esterified to hydroxyl groups of glucose or sucrose molecules. In *S. pennellii*, acylsugars are  
23 mostly 2,3,4-tri-*O*-acylated glucose esters, and some sucrose esters, with C4-C12 fatty acids  
24 (Walters and Steffens, 1990; Shapiro et al., 1994; Schillmiller et al., 2015). Predominant  
25 branched-chain fatty acids (BCFAs) include 2-methylpropanoic acid, 3-methylbutanoic acid, 2-  
26 methylbutanoic acid, and 8-methylnonanoic acid, whereas predominant straight-chain fatty acids  
27 (SCFAs) include *n*-decanoic acid and *n*-dodecanoic acid. Feeding studies demonstrated that  
28 BCFAs are derived from branched-chain amino acids (BCAAs), whereas SCFAs are  
29 hypothesized to be derived by fatty acid synthase (FAS)-mediated de novo fatty acid  
30 biosynthetic process (Walters and Steffens, 1990).

1 Biosynthesis of acylsugars can be broadly divided into two phases: 1) synthesis of fatty  
2 acyl chains, and 2) esterification of these acyl molecules to glucose or sucrose (Figure 1A).  
3 During acylglucose biosynthesis in *S. pennellii*, free fatty acids are initially esterified to glucose  
4 by UDP-glucose:fatty acid glucosyltransferase (UDP-Glc:FA GT) to create 1-*O*-acylglucose  
5 molecules (Ghangas and Steffens, 1993; Kuai et al., 1997). In the next step, which is catalyzed  
6 by a serine carboxypeptidase-like glucose acyltransferase (SCPL GAT), two molecules of 1-*O*-  
7 acylglucose participate in a disproportionation reaction to generate one molecule of 1,2-di-*O*-  
8 acylglucose and one molecule of glucose (Li et al., 1999; Li and Steffens, 2000). Synthesis of the  
9 final product, 2,3,4-tri-*O*-acylglucose, from 1,2-di-*O*-acylglucose requires removal of acyl chains  
10 from C1 position, and esterification or transfer of acyl chains to C3 and C4 positions of glucose.  
11 Since no tri- or tetra-acylglucose molecules are produced by the SCPL GAT, it is likely that  
12 enzymes belonging to hydrolase and/or transferase families are involved in the final steps of  
13 2,3,4-tri-*O*-acylglucose biosynthesis (Figure 1A). In acylsucrose metabolism, acylsugar  
14 hydrolases (ASHs) and acylsugar acyltransferases (ASATs) catalyze the removal and transfer of  
15 acyl chains, respectively, (Schilmiller et al., 2012; Schilmiller et al., 2015; Fan et al., 2016;  
16 Schilmiller et al., 2016).

17 Acylsugars are exuded through trichomes, predominantly by type IV trichomes in *S.*  
18 *pennellii* (Fobes et al., 1985; Slocombe et al., 2008), and acylsucrose biosynthetic genes are  
19 expressed in trichome tip cells (Schilmiller et al., 2012; Schilmiller et al., 2015; Fan et al., 2016).  
20 However, analysis of periclinal chimeras indicates that acylglucose biosynthesis in *S. pennellii* is  
21 not trichome specific (Kuai et al., 1997). Therefore, we used leaf samples for comparative  
22 transcriptomics studies reported here. Although this approach underrepresents actual expression  
23 differences of trichome-specific genes, it ensures that a complete snapshot of candidate gene  
24 networks could be captured.

25 Different accessions of *S. pennellii* produce different amounts of total acylsugars  
26 (Shapiro et al., 1994). Certain accessions of *S. pennellii*, for example LA0716, can produce  
27 acylsugars up to 20% of its leaf dry weight (Fobes et al., 1985), whereas in many low-acylsugar-  
28 producing accessions, acylsugars make up less than 1% of leaf dry weight (Shapiro et al., 1994).  
29 Here, using comparative transcriptomics and gene co-expression network analysis, we exploited  
30 this natural genetic variation among *S. pennellii* accessions to identify candidate biosynthetic and

1 regulatory genes for this trait. Comparative transcriptomics following inhibition of BCAA  
2 biosynthesis further refined the list of candidate genes and shed light on possible regulatory  
3 mechanisms of acylsugar production. Comparative transcriptomics and co-expression studies  
4 also revealed active genetic networks involved in two distinct defense pathways.

5

## 6 **RESULTS**

### 7 **Comparative transcriptomics between low- and high-acylsugar-producing accessions**

8 We performed differential gene expression analysis between a group of three low-acylsugar-  
9 producing accessions (LA1911, LA1912, and LA1926; collectively referred to as ‘LOW’ in this  
10 paper) and a group of three high-acylsugar-producing accessions (LA1941, LA1946, and  
11 LA0716; collectively referred to as ‘HIGH’ in this paper) of *S. pennellii* (Supplemental Table 1)  
12 (Shapiro et al., 1994). Three individual plants for each of the six accessions, with an average of  
13 >25 million processed paired-end reads for each plant, were used for this analysis (Supplemental  
14 Table 2). 19,379 genes passed our filtering criteria for minimum expression levels, and we  
15 obtained 1679 differentially expressed genes (DEGs; see Methods) (Supplemental Dataset 1:  
16 Sheet 1). Of the 1679 DEGs, 931 were upregulated and 748 were downregulated in the ‘HIGH’  
17 group.

18 We also compared transcriptomes between another low-acylsugar-producing accession,  
19 LA1920, and the high-acylsugar-producing accession LA0716, with each accession having four  
20 biological replicates and slightly higher sequencing coverage, as an independent examination of  
21 differential gene expression. In this analysis, 19,967 genes passed our filtering criteria for  
22 minimum expression levels (see Methods), and we identified 3524 DEGs (1465 upregulated and  
23 2059 downregulated genes in LA0716) (Supplemental Dataset 1: Sheet 2). To identify candidate  
24 genes most likely to be involved in acylsugar metabolism, we compared ‘LOW’ vs. ‘HIGH’  
25 DEGs with LA1920 vs. LA0716 DEGs (Figure 1B). 1087 DEGs were common to both  
26 comparisons (586 upregulated and 501 downregulated genes in high-acylsugar-producing  
27 accessions; Supplemental Dataset 1: Sheet 3). Analysis of Gene Ontology (GO) terms associated  
28 with the annotated DEGs revealed that genes involved in fatty acid metabolism, regulation of  
29 transcription and translation, as well as genes whose products are predicted to localize to

1 chloroplasts (the location for BCAA and SCFA biosynthesis), were strongly represented in the  
2 1087 DEGs (Supplemental Figure 1).

3 We also examined the leaf transcriptome of accession LA1302 (three biological  
4 replicates), which accumulates an intermediate level of acylsugars (referred to as ‘MEDIUM’ in  
5 this paper; Supplemental Table 1) (Shapiro et al., 1994), to determinate whether expression  
6 levels of candidate genes were consistent with the amount of acylsugar production (Figure 1B).  
7 Multi-dimensional scaling plots showed a clear distinction in gene expression profiles among  
8 groups of accessions that accumulate different amounts of acylsugars (Supplemental Figure 2A  
9 and 2B). Volcano plots showed high  $\log_2$ fold-change values associated with low false discovery  
10 rate (FDR) values, indicating robust differential expression (Supplemental Figure 2C and 2D).  
11 On the other hand, ten ‘housekeeping genes’ showed similar expression levels across all  
12 biological groups (Supplemental Figure 3).

13

#### 14 **Branched-chain amino acid (BCAA) metabolic genes are upregulated in high-acylsugar-** 15 **producing accessions**

16 Because BCAAs are precursors to acylsugar BCFAs (Walters and Steffens, 1990), we examined  
17 genes involved in BCAA metabolism. Many genes involved in both biosynthesis of BCAAs in  
18 plastids and conversion of BCAAs to branched-chain acyl molecules in mitochondria (Binder,  
19 2010; Maloney et al., 2010) were upregulated in high-acylsugar-producing accessions (Table 1;  
20 Supplemental Figure 4). Most of these BCAA/BCFA metabolic DEGs showed intermediate  
21 levels of expression in the ‘MEDIUM’ accession LA1302 (Figure 2A). Correlation analysis  
22 (Spearman’s rank correlation coefficient, SRCC; denoted by  $\rho$ ) using gene expression profiles in  
23 our 29 samples revealed strong positive correlation among most of these genes (Figure 3).

24

#### 25 **Genes putatively encoding a KAS IV/ KAS II-like enzyme and other FAS components are** 26 **upregulated in high-acylsugar-producing accessions**

27 *Sopen12g004230*, predicted to encode a beta-ketoacyl-(acyl-carrier-protein) synthase II (KAS  
28 II)-like enzyme, had 39-fold higher expression in the ‘HIGH’ group (FDR= 4.40E-18). The

1 SCFA profile of *S. pennellii* acylsugars (predominantly C10 and C12, with a trace amount of C8  
2 in LA1302) is strikingly similar to that of the seed oil in many *Cuphea* species (Shapiro et al.,  
3 1994; Dehesh et al., 1998; Slabaugh et al., 1998; Schutt et al., 2002). A specialized version of  
4 KAS II, referred to as KAS IV, is an important determinant of chain length in *Cuphea* medium-  
5 chain (C8-C12) fatty acids (Dehesh et al., 1998; Slabaugh et al., 1998; Schutt et al., 2002). We  
6 compared amino acid sequences of five KAS IV/ KAS II-like enzymes from four *Cuphea* species  
7 with the Sopen12g004230 sequence, but no similarity was found. However, sequence analysis of  
8 the adjacent Sopen12g004240 coding region (1785-aa; annotated as reverse transcriptase)  
9 revealed a KAS domain at the C-terminal end (1374-1784), and BLAST analysis of *Cuphea*  
10 KAS IV/ KAS II-like enzymes showed high sequence similarity with Sopen12g004240 (77%-  
11 83% identity, 87%-92% similarity, 82%-90% coverage, e-value 0.0; Supplemental Table 3).  
12 Apparently, a transposon inserted into an intron of *Sopen12g004230-Sopen12g004240*, which  
13 led to misannotation of this locus as two separate genes. We confirmed that this locus produces a  
14 single transcript with RT-PCR analysis (Supplemental Figure 5). Blauth et al. (1999) used an  
15 intraspecific F<sub>2</sub> population derived from the cross between accessions LA0716 and LA1912, to  
16 identify a quantitative trait locus on the upper arm of chromosome 12 that is associated with C10  
17 SCFA production (Blauth et al., 1999). The locus containing *Sopen12g004230-Sopen12g004240*  
18 is in this region.

19 DEGs putatively encoding other components of the FAS complex were also upregulated  
20 in high-acylsugar-producing accessions (Table 1). Expression patterns of these genes and  
21 *Sopen12g004240* (KAS IV/ KAS II-like) showed very strong positive correlation among  
22 themselves, as well as with BCAA metabolic DEGs (Figure 3; Supplemental Dataset 2).

23

#### 24 **Expression patterns of four genes putatively encoding acyl-activating enzymes show very** 25 **strong positive correlation with FAS components**

26 Acyl-activating enzymes (AAEs) are ATP/AMP-binding proteins that activate different  
27 carboxylic acids by forming fatty acyl-CoA molecules from free fatty acids, ATP and CoA  
28 (Shockey et al., 2003). We found four DEGs putatively encoding acyl-activating enzyme 1  
29 (AAE1), and they had 8- to 56-fold higher expression in the 'HIGH' group (FDR= 2.80E-07 to

1 3.80E-27; Table 1). In *Arabidopsis thaliana*, AAE7 and AAE11, which belong to the same  
2 phylogenetic clade as AAE1, show selective activity against short- and medium-chain fatty acids  
3 (Shockey et al., 2003); interestingly, these four AAE1 genes exhibited very strong positive  
4 correlation with FAS components ( $\rho = 0.87$  to  $0.98$ ;  $P = 1.07E-9$  to  $4.44E-16$ ; Figure 3).

5

### 6 **Acylsucrose metabolic genes are upregulated in high-acylsugar-producing accessions and** 7 **show strong positive correlation with phase 1 DEGs**

8 Three genes encoding acylsugar acyltransferases (ASATs), which are involved in phase 2 of  
9 acylsucrose biosynthesis (Schillmiller et al., 2015; Fan et al., 2016; Fan et al., 2017), were  
10 upregulated in high-acylsugar-producing accessions (Table 2). ASATs use acyl-CoA molecules  
11 as their substrates (Schillmiller et al., 2012; Schillmiller et al., 2015; Fan et al., 2016; Fan et al.,  
12 2017), and free straight-chain fatty acids (SCFA) produced by FAS-catalyzed de novo fatty acid  
13 biosynthesis must be activated to their acyl-CoA derivatives (SCFA-CoA) before they can be  
14 used by ASATs (Figure 1A). Interestingly, all three ASAT genes exhibited very strong positive  
15 correlation with both AAE1 genes and FAS components (Figure 3; Supplemental Dataset 2).  
16 Using data of Ning et al. (2015), who published expression profiles of *S. lycopersicum* genes in  
17 isolated stem trichomes versus shaved stems, we also investigated if AAE1 homologs in *S.*  
18 *lycopersicum* show trichome-enriched expression pattern, since ASAT genes are expressed in  
19 apical cells of type I/IV glandular trichomes (Schillmiller et al., 2012; Schillmiller et al., 2015;  
20 Fan et al., 2016). We identified *Solyc02g082870*, *Solyc02g082880*, *Solyc07g043630*, and  
21 *Solyc07g043660* as putative orthologs of the four AAE1 DEGs *Sopen02g027670*,  
22 *Sopen02g027680*, *Sopen07g023200*, and *Sopen07g023220*, respectively (Supplemental Dataset  
23 3). According to Ning et al. (2015), these four tomato genes show 300-, 26-, 562-, and 294-fold  
24 higher expression levels, respectively, in isolated stem trichomes compared to shaved stems. In  
25 fact, three of these AAE1 homologs show almost no expression in shaved stems, but high  
26 expression levels in isolated trichomes, similar to ASAT genes.

27 Recently, three genes encoding acylsugar hydrolases (ASHs; carboxylesterases that  
28 remove acyl groups from acylsucrose molecules) and a related carboxylesterase gene  
29 (*Sopen04g001210*) were reported in cultivated and wild tomato (Schillmiller et al., 2016). These



1 genes, except *Sopen05g030120* (Sp-*ASH1*), were in the list of 1087 common DEGs, and showed  
2 strong positive correlation with genes putatively involved in phase 1 acyl chain synthesis (Figure  
3 3; Supplemental Dataset 2). The expression profile of *ASH1* in different tissues of *S.*  
4 *lycopersicum* led the authors to propose additional non-trichome-localized functions for Sl-  
5 *ASH1* (Schillmiller et al., 2016); our results further support the idea that *ASH1* may not have a  
6 critical role in trichome acylsucrose metabolism.

7

### 8 **Expression levels of two acylglucose biosynthetic genes are not consistent with acylsugar** 9 **levels**

10 The first two steps in phase 2 of acylglucose biosynthesis are catalyzed by UDP-Glc:FA GT and  
11 SCPL GAT, respectively (Figure 1A) (Ghangas and Steffens, 1993; Kuai et al., 1997; Li et al.,  
12 1999; Li and Steffens, 2000). Interestingly, *Sopen01g049990*, encoding the UDP-Glc:FA GT,  
13 showed 2-fold higher expression in the ‘LOW’ group (FDR= 2.70E-03), whereas  
14 *Sopen10g020280*, which encodes the SCPL GAT, had no significant difference in expression  
15 between low- and high-acylsugar-producing accessions (Table 2). Similarly, *Sopen01g049990*  
16 showed weak negative correlation with genes putatively involved in acyl chain synthesis (phase  
17 1), whereas *Sopen10g020280* showed no significant correlation (Figure 3; Supplemental Dataset  
18 2).

19 The UDP-Glc:FA GT uses free fatty acids (such as 2-methylpropanoate), but not their  
20 CoA derivatives, to initiate acylglucose biosynthesis (Ghangas and Steffens, 1993). In  
21 mitochondria, activity of the branched-chain keto acid dehydrogenase complex forms BCFA-  
22 CoA derivatives of BCAAs, and these acyl-CoA molecules therefore must be released as free  
23 fatty acids (BCFAs) by thioesterases before they could be used for acylglucose biosynthesis  
24 (Figure 1A). One gene (*Sopen12g032690*), predicted to encode a mitochondrial acyl-CoA  
25 thioesterase, had 4-fold higher expression in the ‘HIGH’ group (FDR= 7.80E-03; Table 1).

26

### 27 **Three genes putatively encoding ATP-binding cassette (ABC) transporters are upregulated** 28 **in high-acylsugar-producing accessions**

1 One important question in acylsugar metabolism is how acylsugar molecules are exported out of  
2 trichome cells. It remains unclear whether acylsugars are transported across the plasma  
3 membrane via membrane-associated transporter proteins or via vesicular transport. Five genes  
4 belonging to the ATP-binding cassette (ABC) transporter family showed differential expression,  
5 and three of them (*Sopen12g034820*, *Sopen04g005380*, and *Sopen03g001870*) were upregulated  
6 in high-acylsugar-producing accessions (Table 3). We identified *Solyc12g100190*,  
7 *Solyc04g010200*, and *Solyc03g005860* as putative orthologs of these three ABC transporters  
8 (Supplemental Dataset 3). Based on data of Ning et al. (2015), these three tomato genes show  
9 143-, 160-, and 470-fold, respectively higher expression in isolated trichomes than in underlying  
10 tissues, with expression of *Solyc03g005860* being very similar to that of trichome-specific ASAT  
11 genes. Because acylsugars are secreted by trichome tip cells (Fobes et al., 1985), it is reasonable  
12 to expect that genes involved in acylsugar secretion will exhibit trichome-specific/ trichome-  
13 enriched expression pattern. On the other hand, putative *S. lycopersicum* orthologs  
14 (*Solyc04g051800* and *Solyc01g105450*; Supplemental Dataset 3) of the other ABC transporter  
15 DEGs (*Sopen04g023150* and *Sopen01g047950*), which were downregulated in high-acylsugar-  
16 producing accessions, have similar expression levels in both trichomes and underlying tissues  
17 (Ning et al., 2015).

18

## 19 **Genes involved in central carbon metabolism are upregulated in high-acylsugar-producing** 20 **accessions**

21 In high-acylsugar-producing accessions, acylsugar accumulation represents an astounding  
22 metabolic investment (up to 20% of leaf dry weight in LA0716) (Fobes et al., 1985; Shapiro et  
23 al., 1994), and we investigated whether genes involved in central carbon metabolism were  
24 differentially expressed to support acylsugar production. Two genes (*Sopen07g006810* and  
25 *Sopen08g020180*) putatively encoding the RUBISCO small subunit and chloroplast NADP-  
26 malic enzyme, respectively were upregulated in high-acylsugar-producing accessions (Table 3).  
27 Compared to other RUBISCO small subunit genes, which were not differentially expressed and  
28 had very high expression levels (FPKM as high as 5000; Supplemental Dataset 1), the DEG  
29 *Sopen07g006810* had moderate levels of expression in medium- and high-acylsugar-producing  
30 accessions, and very low levels of expression in low-acylsugar-producing accessions (Figure

1 2C). Putative orthologs of *Sopen07g006810* and *Sopen08g020180* in *S. lycopersicum*  
2 (*Solyc07g017950* and *Solyc08g066360*, respectively; Supplemental Dataset 3) show 47- and 16-  
3 fold higher expression, respectively, in isolated stem trichomes compared to shaved stems (Ning  
4 et al., 2015).

5 Two other genes (*Sopen03g040490* and *Sopen01g042050*) annotated as ‘invertase’ and  
6 ‘sugar transporter ERD6-like isoform 5’, respectively, were also upregulated in high-acylsugar-  
7 producing accessions (Table 3). ERD6-like transporters are capable of monosaccharide uptake  
8 by facilitated diffusion (Yamada et al., 2010). The putative ortholog of *Sopen03g040490*  
9 (invertase) in *S. lycopersicum* (*Solyc03g121680*) shows no expression in isolated trichomes,  
10 whereas *Solyc01g098490* (putative ortholog of the monosaccharide transporter *Sopen01g042050*;  
11 Supplemental Dataset 3) shows 22-fold higher expression in trichomes compared to non-  
12 trichome tissues (Ning et al., 2015).

13

#### 14 **Genes putatively encoding oxylipin metabolic proteins and other defense proteins are** 15 **upregulated in low-acylsugar-producing accessions**

16 Because acylsugars have important roles in plant defense (Hawthorne et al., 1992; Juvik et al.,  
17 1994; Liedl et al., 1995; Chortyk et al., 1997; Hare, 2005; Luu et al., 2017), we investigated if  
18 defense genes were differentially expressed between low- and high- acylsugar-producing  
19 accessions. Nine genes putatively encoding fatty acid desaturase and lipoxygenase enzymes were  
20 differentially expressed, and eight of them were upregulated in low-acylsugar-producing  
21 accessions (Table 4). Lipoxygenases catalyze the dioxygenation of polyunsaturated fatty acids  
22 produced by fatty acid desaturases, and this enzymatic conversion is the first step in the  
23 biosynthesis of oxylipins, which contribute to innate immunity (La Camera et al., 2004).  
24 Oxylipin response pathways are regulated by TGA transcription factors, and two classes of  
25 enzyme (oxo-phytodienoate reductase and glutathione S-transferase) have been reported to take  
26 important roles in metabolizing highly reactive oxylipins into less reactive molecules (Mueller et  
27 al., 2008). Genes putatively encoding these oxylipin signaling and metabolic proteins were also  
28 upregulated in low-acylsugar-producing accessions (Table 4).

1           Lipoxygenase-produced oxylipins have important roles in plant defense against  
2 herbivores (Halitschke and Baldwin, 2003) and microbial pathogens (Prost et al., 2005).  
3 Therefore, low-acylsugar-producing accessions may generate oxylipins to compensate for  
4 diminished antiherbivory and antimicrobial activities of acylsugars. Consistent with this  
5 hypothesis, genes putatively encoding subtilisin-like proteases, which play important role in the  
6 recognition of plant pathogens and activation of immune responses (Figueiredo et al., 2014), and  
7 additional proteins involved in plant defense response to fungal and other biotic stimulus were  
8 also upregulated in low-acylsugar-producing accessions (Table 4). Based on expression profiles,  
9 genes involved in fatty acid desaturation, lipoxygenation, oxylipin signaling and metabolism, and  
10 antimicrobial defense showed moderate to strong correlation with each other (Supplemental  
11 Figure 6), indicating an active network of innate immunity in low-acylsugar-producing  
12 accessions.

13

#### 14 **Amine oxidase-related defense response genes are upregulated in high-acylsugar-producing** 15 **accessions**

16 We found that one DEG (*Sopen06g005510*) predicted to encode the catalytic domain of  
17 glutamine synthetase had 183-fold higher expression in the 'HIGH' group than in the 'LOW'  
18 group (FDR= 2.60E-42). In plants, expression of glutamine synthetase is induced by higher  
19 concentrations of ammonia, which encouraged us to search for DEGs involved in ammonia-  
20 liberating reactions. Three genes encoding amine oxidase domains (*Sopen07g017370*,  
21 *Sopen11g004040*, and *Sopen11g004050*) and an aldehyde oxidase gene (*Sopen01g035050*),  
22 whose product could potentially oxidize aldehydes produced by the amine oxidases, were also  
23 noticeably upregulated in high-acylsugar-producing accessions (Table 4). BLAST results  
24 indicate that *Sopen11g004040* and *Sopen11g004050* show similarity with FLOWERING  
25 LOCUS D-like proteins, which have a flavin-containing amine oxidase domain and are required  
26 for systemic acquired resistance in *Arabidopsis thaliana* (Singh et al., 2013).

27           H<sub>2</sub>O<sub>2</sub> generated by amine oxidase- and aldehyde oxidase-catalyzed reactions acts as a  
28 signal molecule and plays an important role during pathogen invasion by acting on extensin  
29 proteins (Niebel et al., 1993; Jackson et al., 2001; Cona et al., 2006). Two genes predicted to

1 encode extensin-like glycoproteins were upregulated in high-acylsugar-producing accessions  
2 (Table 4), and one of them (*Sopen05g003410*) showed strong positive correlation with amine  
3 oxidase- and aldehyde oxidase-encoding genes (Supplemental Figure 6). Although amine  
4 oxidase-related defense genes were upregulated in high-acylsugar-producing accessions, it is  
5 important to note that their expression levels were much lower than some oxylipin-related  
6 defense genes in low-acylsugar-producing accessions (Figure 4).

7

## 8 **Identification of co-expressed gene networks associated with acylsugar metabolism and** 9 **plant defense**

10 Strong correlation among expression profiles of known and putative acylsugar metabolic genes  
11 (Figure 3) encouraged us to perform gene co-expression network analysis. Using weighted gene  
12 correlation network analysis (WGCNA) (Langfelder and Horvath, 2008), we identified 15  
13 modules (clusters of co-expressed genes). As a representative of gene expression profile in each  
14 WGCNA module, a module eigengene was defined as the first principle component in each  
15 module (Langfelder and Horvath, 2008). Acylsugar accumulation and ASAT gene expression  
16 showed strong positive correlation with module eigengene of the ‘turquoise’ module, and strong  
17 negative correlation with module eigengene of the ‘black’ module (Figure 5). Of the 586  
18 upregulated genes in high-acylsugar-producing accessions, 545 (93%) were found in the  
19 ‘turquoise’ module, whereas 371 (74%) of the 501 downregulated genes were clustered in the  
20 ‘black’ module (Supplemental Dataset 4: Sheet 1). The top-100 genes most strongly correlated  
21 with acylsugar accumulation in the ‘turquoise’ module included genes involved in BCAA/BCFA  
22 metabolism, SCFA metabolism (putative FAS components and acyl-activating enzymes AAE1),  
23 acylsucrose metabolism (except *Sp-ASH1*), transport process (ABC transporters), and central  
24 carbon metabolism (Supplemental Dataset 4: Sheet 3). On the other hand, the ‘black’ module  
25 contained genes for putative oxylipin-mediated defense pathway, and two  
26 carboxylesterase/acylhydrolase genes (*Sopen02g020660* and *Sopen02g020670*) that exhibited  
27 strong negative correlation with acylsugar accumulation ( $\rho = -0.79$  and  $-0.85$ , respectively;  $P =$   
28  $4.56E-07$  and  $3.84E-09$ , respectively).

1 We also selected *Sopen01g002520* (lipoxygenase) and *Sopen11g004040* (FLOWERING  
2 LOCUS D) as representative of the oxylipin-mediated and the amine oxidase-mediated defense  
3 networks, respectively, and identified genes that are strongly correlated with them. Genes  
4 putatively encoding fatty acid desaturase, TGA transcription factor, glutathione S-transferase,  
5 subtilisin-like protease and other pathogenesis-related proteins were associated with the  
6 lipoxygenase gene (Supplemental Dataset 4: Sheet 7). On the other hand, a search for top-100  
7 genes associated with *Sopen11g004040* in the ‘turquoise’ module identified the adjacent  
8 FLOWERING LOCUS D gene *Sopen11g004050*, as well as genes predicted to encode copper-  
9 containing amine oxidase, aldehyde oxidase, extensin-like protein, leucine rich repeat (LRR)  
10 containing serine/threonine protein kinase, tyrosine protein kinase, NB-ARC domain-containing  
11 protein, exostosin family glycosyltransferases, F-box proteins, transcription factors  
12 (Supplemental Table 4), and more than 50 hypothetical proteins (Supplemental Dataset 4: Sheet  
13 8).

14 It is interesting to note that a protamine P1-like protein (*Sopen09g015380*; 154-aa; 24  
15 Lys, 29 Arg) and some hypothetical proteins encoded by the genes strongly correlated with  
16 *Sopen11g004040*, had higher content of Lys and/or Arg residues (Supplemental Figure 7). The  
17 biological implication of this observation is not clear; however, small, cationic peptides from  
18 other plants have antimicrobial properties (Marcos et al., 2008), and co-expression of these genes  
19 with *Sopen11g004040* may reflect another branch of the plant defense network.

20

## 21 **Protein kinases are correlated with the UDP-Glc:FA GT, whereas factors involved in** 22 **translation are strongly associated with the SCPL GAT**

23 Because expression profiles of the two phase 2 acylglucose biosynthetic genes were not  
24 consistent with acylsugar accumulation (Table 2), we attempted to find and analyze genes that  
25 are strongly associated with them. The expression profile of *Sopen01g049990* (UDP-Glc:FA GT)  
26 showed strong correlation with module eigengenes in the ‘lightcyan’ and ‘tan’ modules, whereas  
27 *Sopen10g020280* (SCPL GAT) was associated with the ‘green’ module (Figure 5). We selected  
28 top-100 genes associated with each of *Sopen01g049990* and *Sopen10g020280* in these modules,  
29 and enrichment analysis revealed that Gene Ontology (GO) terms such as ‘phosphotransferase

1 activity' (GO:0016773;  $P= 3.32E-04$ ), 'protein kinase activity' (GO:0004672;  $P= 5.82E-04$ ), and  
2 'transcription factor activity' (GO:0003735;  $P= 3.40E-02$ ) were overrepresented in the  
3 'lightcyan' and 'tan' modules; on the other hand, GO terms such as 'RNA binding'  
4 (GO:0003723;  $P= 3.57E-06$ ) and 'structural constituent of ribosome' (GO:0003735;  $P= 1.96E-$   
5  $05$ ) were significantly overrepresented in the 'green' module (Figure 6; Supplemental Figure 8).  
6 In fact, 11 genes encoding ribosomal proteins, as well as many genes involved in post-  
7 transcriptional processing and translation process were in the list of top-100 genes associated  
8 with *Sopen10g020280* (Supplemental Table 5). Interestingly, this list also included two genes  
9 encoding signal recognition particles, consistent with a previous report that the SCPL GAT is a  
10 glycoprotein with a cleavable N-terminal 18-aa signal peptide (Li and Steffens, 2000).

11 Because *Sopen10g020280* (SCPL GAT) does not show significant differential expression  
12 between low- and high-acylsugar-producing accessions, we considered the possibility that genes  
13 associated with it in the 'green' module were simply housekeeping genes whose expression did  
14 not vary, either. However, 9 of our 10 selected housekeeping genes, as well as many genes  
15 putatively encoding signal recognition particles and ribosomal proteins, showed little or no  
16 correlation with *Sopen10g020280* and were placed in other modules (Supplemental Table 6).  
17 These results suggest that some genes in the 'green' module, whose expression profiles are  
18 strongly correlated with that of *Sopen10g020280*, may have a role in regulating the SCPL GAT  
19 activity at the level of translation.

20

## 21 **Inhibition of BCAA biosynthesis changes transcript levels of many acylsugar candidate** 22 **DEGs, but not defense DEGs**

23 Acetolactate synthase (ALS) catalyzes the first common step in the biosynthesis of BCAAs  
24 (Supplemental Figure 4A), and inhibition of this enzyme significantly lowers acylsugar  
25 production in *S. pennellii* (Walters and Steffens, 1990). To determine whether inhibition of this  
26 enzyme changed expression of ALS and other candidate genes involved in acylsugar  
27 metabolism, we treated leaves of the high-acylsugar-producing accession LA0716 with the ALS  
28 inhibitor imazapyr at 0.1 mM and 1 mM concentrations. Many of the DEGs involved in  
29 BCAA/BCFA metabolism showed differential expression in response to imazapyr treatment

1 (Table 5). Interestingly, at 1 mM imazapyr, expression of *Sopen04g026270* and  
2 *Sopen01g028100* (branched-chain keto acid dehydrogenase complex E1 and E2 subunits,  
3 respectively) was significantly increased, possibly because the amount of 2-ketobutyrate, a  
4 substrate of ALS that also can be used directly by the dehydrogenase complex to make  
5 propionyl-CoA, builds up (Walters and Steffens, 1990).

6 Surprisingly, DEGs putatively involved in SCFA metabolism (FAS components and  
7 AAE1 members), showed significant downregulation at both 0.1 mM and 1 mM imazapyr in a  
8 concentration-dependent manner. It is interesting to note that AAE1 proteins have a peroxisomal  
9 location in *Arabidopsis thaliana* (Reumann et al., 2009); one gene (*Sopen11g007710*) predicted  
10 to encode Mpv17/PMP22 family peroxisomal membrane protein had 7.7-fold (FDR= 1.17E-08)  
11 and 165-fold (FDR= 3.62E-24) lower expression at 0.1 mM and 1 mM imazapyr treatment,  
12 respectively (Supplemental Dataset 5). *Sopen11g007710* was upregulated 29-fold in the ‘HIGH’  
13 group (FDR= 2.80E-11; Supplemental Dataset 1), and its putative ortholog in *S. lycopersicum*  
14 (*Solyc11g012990*; Supplemental Dataset 3) has 1420-fold higher expression in isolated  
15 trichomes than shaved stems, consistent with trichome-enriched expression of AAE1 members  
16 (Ning et al., 2015).

17 *Sopen01g049990* (UDP-Glc:FA GT; phase 2 of acylglucose biosynthesis) showed 3.5-  
18 fold higher expression at both concentrations of imazapyr (FDR= 2.59E-06 and 4.01E-07,  
19 respectively), supporting our previous finding that a negative correlation may exist between  
20 acylsugar level and expression level of this gene. *Sopen10g020280* (SCPL GAT) did not show  
21 any significant change in expression level at either concentration ( $\log_2FC > 1$ , FDR < 0.05). On  
22 the other hand, DEGs encoding ASATs (phase 2 of acylsucrose biosynthesis) showed significant  
23 decrease in gene expression level in a concentration-dependent manner (by as much as 94-fold,  
24 FDR= 1.65E-14), as did ASHs and the related carboxylesterase *Sopen04g001210* (by as much as  
25 70-fold, FDR= 8.82E-50) (Table 5).

26 Three DEGs putatively encoding ABC transporters were significantly downregulated at  
27 both concentrations of imazapyr (by as much as 32-fold, FDR= 4.43E-14). Imazapyr treatment in  
28 *Arabidopsis thaliana* resulted in significant induction of nine genes (out of the ten genes  
29 reported) encoding ABC transporters, indicating their role in detoxification process (Manabe et  
30 al., 2007). Similarly, 15 putative ABC transporter genes in *S. pennellii* were upregulated in



1 response to imazapyr either at higher concentration or at both concentrations (Supplemental  
2 Table 7). However, three putative ABC transporter DEGs were repressed by imazapyr treatment,  
3 consistent with their inclusion in the WGCNA ‘turquoise’ module with other acylsugar  
4 metabolic genes.

5 Interestingly, DEGs involved in central carbon metabolism were also significantly  
6 downregulated in response to imazapyr (by as much as 45-fold, FDR= 4.99E-10 for RUBISCO  
7 small subunit; Table 5). While investigating the role of central carbon metabolism in specialized  
8 metabolism of *S. lycopersicum* and *S. habrochaites* glandular trichomes, Balcke et al. (2017)  
9 reported importance of glutathione metabolism to cope with oxidative stress associated with high  
10 productivity of specialized metabolites in trichomes (Balcke et al., 2017). One glutathione S-  
11 transferase gene (*Sopen12g026980*) was upregulated in high-acylsugar-producing accessions  
12 (Figure 4) and was placed in the ‘turquoise’ module due to strong positive correlation between  
13 expression profile of this gene and acylsugar accumulation ( $\rho= 0.82$ ;  $P= 5.93E-08$ ).  
14 *Sopen12g026980* showed 2- and 18-fold (FDR= 2.36E-02 and 8.94E-18, respectively)  
15 downregulation at 0.1 mM and 1 mM imazapyr, respectively (Supplemental Dataset 5).  
16 According to data of Ning et al. (2015), the putative ortholog of this gene in *S. lycopersicum*  
17 (*Solyc05g006730*) shows 287-fold higher expression in isolated stem trichomes compared to  
18 underlying tissues.

19 The 14 genes strongly associated with the FLOWERING LOCUS D gene  
20 *Sopen11g004040* (Supplemental Table 4) did not show any significant differential expression  
21 ( $\log_2FC > 1$ , FDR < 0.05) in response to imazapyr treatment at either concentration, except two  
22 genes which showed slight changes in gene expression only at one concentration (Supplemental  
23 Table 8). On the other hand, DEGs with known and putative functions in acylsugar metabolism  
24 responded considerably to ALS inhibition. These results suggested that additional genes involved  
25 in acylsugar metabolism may also be represented in the list of 171 DEGs (out of 1087 common  
26 DEGs between low- and high-acylsugar-producing accessions) that responded to imazapyr at  
27 both concentrations (Supplemental Dataset 5). In this list, 13 genes putatively encoding  
28 transcription factors were present; of these, nine had higher expression levels in high-acylsugar-  
29 producing accessions (Table 3), and eight of these were found in the WGCNA ‘turquoise’  
30 module.

1

2 **Three AP2-family transcription factor genes form a strong co-expression network with**  
3 **many acylsugar and flavonoid metabolic genes**

4 We selected the top-50 most strongly connected genes (based on intramodular connectivity as  
5 determined by WGCNA) for each of the eight above-mentioned transcription factor genes, and  
6 merged them to identify the most strongly interconnected gene network. Three genes  
7 (*Sopen12g021250*, *Sopen03g036630*, and *Sopen05g008450*) putatively encoding AP2-family  
8 transcription factors and one gene (*Sopen10g031080*) encoding a homeobox-associated leucine-  
9 zipper protein revealed strongest intramodular connectivity with acylsugar metabolic genes  
10 (Figure 7). Astonishingly, 48 genes were common among the top-50 most strongly connected  
11 genes for each of the three AP2-family members (Supplemental Figure 9).

12 This network included genes involved in BCFA synthesis, SCFA synthesis (FAS  
13 components) and their activation (AAEs and the related peroxisomal membrane protein  
14 Mpv17/PMP22), acylsucrose metabolism (three ASATs, ASH3 and the related carboxylesterase  
15 *Sopen04g001210*), as well as genes putatively encoding three ABC transporters, RUBISCO  
16 small subunit, glutathione S-transferase, and other proteins. According to data of Ning et al.  
17 (2015), putative orthologs of the three AP2-family transcription factor genes in *S. lycopersicum*  
18 (*Solyc12g042210*, *Solyc03g117720*, and *Solyc05g013540*, respectively; Supplemental Dataset 3)  
19 show more than 205-, 24-, and 252-fold higher expression in isolated trichomes compared to  
20 shaved stems, whereas *Solyc10g080540* (putative ortholog of *Sopen10g031080*; leucine-zipper  
21 protein) does not show any significant difference in expression between isolated trichomes and  
22 shaved stems (1.2-fold, FDR= 1).

23 However, not all genes in this network had an obvious connection to acylsugar  
24 metabolism. One gene on chromosome 11 (*Sopen11g003320*; UDP-glucose:catechin  
25 glucosyltransferase (Noguchi et al., 2008)) and three sequential genes on chromosome 6  
26 (*Sopen06g034810*, *Sopen06g034820*, and *Sopen06g034830*; myricetin methyltransferase  
27 (Schmidt et al., 2012; Kim et al., 2014)) showed strong intramodular connectivity with AP2-  
28 family transcription factors and other acylsugar metabolic genes (Figure 7). These flavonoid  
29 metabolic genes were also strongly downregulated in response to imazapyr treatment

1 (Supplemental Figure 10), and their putative orthologs in *S. lycopersicum* are preferentially  
2 expressed in trichomes (Ning et al., 2015). In addition to abundant acylsugar compounds,  
3 presence of flavonoid compounds, such as methylated myricetin, has been reported in *S.*  
4 *pennellii* (LA0716) trichomes (McDowell et al., 2011). These results suggest that a subset of  
5 AP2-family transcription factors is strongly connected with metabolism of both acylsugar and  
6 flavonoid compounds.

7

## 8 **DISCUSSION**

9 Acylsugars are powerful natural pesticides (Puterka et al., 2003), and increasing acylsugar-  
10 mediated insect pest resistance has been a target of various breeding programs in solanaceous  
11 crops (Bonierbale et al., 1994; Lawson et al., 1997; Alba et al., 2009; Leckie et al., 2012). The  
12 success of these breeding programs and bioengineering of acylsugar production will largely  
13 depend on unraveling the complex network of structural and regulatory genes. Here, we  
14 exploited natural variation in acylsugar production among *S. pennellii* accessions through a  
15 combination of comparative transcriptomics study and gene co-expression network analysis to  
16 identify a number of genes with known and putative functions in acylsugar metabolism. Our  
17 initial hypothesis for this work was that expression levels of known acylsugar biosynthetic genes  
18 would be consistent with acylsugar levels in different accessions of *S. pennellii*, and that  
19 additional, unknown genes involved in acylsugar metabolism would be coordinately expressed  
20 with them. Our approach was validated by the identification of several DEGs with known  
21 functions in acylsugar biosynthesis, such as BCAA/BCFA metabolic genes and acylsucrose  
22 biosynthetic genes (ASATs). Our work also revealed differential expression of two defense  
23 networks among *S. pennellii* accessions.

24

## 25 **Defense pathway genes**

26 Higher expression levels of genes putatively encoding oxylipin metabolic proteins, subtilisin-like  
27 proteases, and other defense proteins in low-acylsugar-producing accessions (Table 4) may  
28 indicate a compensatory mechanism for diminished defense activities of acylsugars in these

1 accessions. Alternatively, expression of these defense genes may be suppressed in high-  
2 acylsugar-producing accessions due to the protection provided by acylsugars. Our data do not  
3 allow us to distinguish between these two possibilities, but it is important to note that all plants  
4 used in this study were grown in the same growth chamber, and no insects or symptoms of  
5 disease were apparent on any plant.

6 A different class of defense genes, including those putatively encoding amine oxidases,  
7 aldehyde oxidase, and extensin-like proteins, was upregulated in high-acylsugar-producing  
8 accessions. In general, these amine oxidase-related DEGs had much lower expression levels than  
9 some oxylipin-related defense DEGs in low-acylsugar-producing accessions (Figure 4). These  
10 different levels of expression may indicate that defense DEGs in high-acylsugar-producing  
11 accessions are involved in baseline innate immunity, whereas acylsugars provide the major  
12 defensive barrier. During pathogen or herbivore attack, these amine oxidases may be upregulated  
13 to produce H<sub>2</sub>O<sub>2</sub>, which can activate extensin proteins (Cona et al., 2006). Increased levels of  
14 extensin transcripts upon nematode infection has been reported in tobacco (Niebel et al., 1993),  
15 and deposition of extensin mediated by H<sub>2</sub>O<sub>2</sub>-induced peroxidase during callus culture elicitation  
16 has been reported in *Vitis vinifera* (Jackson et al., 2001).

17 Most defense DEGs (both oxylipin-related and amine oxidase-related) had similar  
18 expression levels in the 'MEDIUM' accession LA1302 and in low-acylsugar-producing  
19 accessions (Figure 4). Multi-dimensional scaling plots also showed that the overall gene  
20 expression profile of LA1302 is similar to that of low-acylsugar-producing accessions, but very  
21 different from that of high-acylsugar-producing accessions (Supplemental Figure 2A and 2B).  
22 These results are consistent with the previous prediction that LA1302 is more closely related to  
23 low-acylsugar-producing accessions than high-acylsugar-producing accessions (Shapiro et al.,  
24 1994). Despite the similarity of defense DEGs expression and the overall transcriptome profile  
25 between LA1302 and low-acylsugar-producing accessions, many DEGs with known and putative  
26 functions in acylsugar metabolism showed intermediate expression levels in LA1302 (Figure 2).  
27 Together, these results suggest that there is no strict linear relationship between acylsugar  
28 accumulation and expression of defense pathways. Further studies are required to fully  
29 understand this interaction.

30

## 1 **Candidate genes involved in acylsugar biosynthesis and export**

2 Strong correlations among expression profiles of selected acylsugar metabolic genes (Figure 3)  
3 encouraged us to perform gene co-expression network analysis, which identified candidate genes  
4 involved in acylsugar metabolism. Roles of candidate genes were further supported by their  
5 response to imazapyr treatment, which caused significant downregulation of most acylsugar  
6 metabolic genes (Table 5), but not defense genes (Supplemental Table 8). Additionally, data of  
7 Ning et al. (2015) indicate that putative *S. lycopersicum* orthologs of most of our candidate genes  
8 are preferentially expressed in trichomes. Roles of some candidate genes in acylsugar  
9 metabolism are discussed below.

10 Acylsugar phase 1: Sequence similarity between Sopen12g004240's KAS domain and *Cuphea*  
11 KAS IV/KAS II-like enzymes, which synthesize medium-chain (C8-C12) SCFAs, suggests a  
12 role of Sopen12g004240 (in collaboration with Sopen08g002520; KAS III) in acylsugar SCFA  
13 synthesis. This is consistent with the previous prediction that KAS enzyme(s) other than KAS I  
14 might be involved in SCFA production in *S. pennellii* acylsugars (Slocombe et al., 2008).  
15 Furthermore, *Sopen12g004240* is located near a previously identified SCFA quantitative trait  
16 locus on the upper arm of chromosome 12 (Blauth et al., 1999).

17 We identified putative medium-chain specific AAE1 members (Shockey et al., 2003) that  
18 may provide SCFA-CoA molecules for ASATs in acylsucrose biosynthesis, based on not only  
19 their strong positive correlation with both FAS components and ASATs (Figure 3), but also  
20 trichome-enriched expression patterns of their putative orthologs in *S. lycopersicum* (Ning et al.,  
21 2015). Duplicated genes often encode proteins with similar but distinct substrate specificity, and  
22 DEGs putatively encoding AAE1 members are located on chromosome 2 and chromosome 7 as  
23 members of similar gene clusters in both *S. pennellii* (Bolger et al., 2014a) and *S. lycopersicum*  
24 (The Tomato Genome Consortium, 2012). It will be interesting to see if different AAE1  
25 members can activate SCFAs with different chain lengths.

26 Putative peroxisomal locations for the four AAE1 members (Reumann et al., 2009) are  
27 consistent with their strong connection with *Sopen11g007710* (Figure 7), which is predicted to  
28 encode an Mpv17/PMP22 family peroxisomal membrane protein. PMP22 is presumably  
29 involved in controlling permeability of peroxisomal membrane (Brosius et al., 2002), and

1 *Sopen11g007710* showed strong response to imazapyr treatment, suggesting an important role of  
2 peroxisome in acylsugar metabolism (Figure 8).

3 Acylsugar phase 2: Because no tri- or tetra-acylglucose products are formed by the SCPL GAT  
4 (Li et al., 1999; Li and Steffens, 2000), synthesis of the final product 2,3,4-tri-*O*-acylglucose  
5 from 1,2-di-*O*-acylglucose requires additional, unidentified enzyme(s) (Figure 1A). If these  
6 additional enzymatic activities are not regulated at the transcriptional level, then comparative  
7 transcriptomics would not be able to identify them. However, transcriptional regulation may be  
8 involved in specific segments of this pathway, and the role of few candidate genes can be  
9 discussed here.

10 Although ASHs remove acyl chains from the sucrose pyranose ring in vitro, several  
11 possible in vivo functions have been proposed for ASH activity, including roles in acylsugar  
12 remodeling, replacing acyl chains, or catalyzing the reverse reaction (adding acyl chains to  
13 acylsugar molecules rather than removing them) (Schillmiller et al., 2016). Genes encoding  
14 ASH3 and the related carboxylesterase *Sopen04g001210* showed strong associations with other  
15 acylsugar metabolic genes (Figure 7) and also responded to imazapyr in a similar way (Table 5).  
16 It is possible that these enzymes catalyze in vivo acylation of 1,2-di-*O*-acylglucose at C3 and C4  
17 positions and/or removal or remodeling of the C1 acyl group to finally synthesize 2,3,4-tri-*O*-  
18 acylglucose. *Sopen04g001210* is of special interest because it has very high expression levels in  
19 our leaf samples (Figure 2B), and its putative ortholog in *S. lycopersicum* shows trichome-  
20 enriched expression (Ning et al., 2015; Schillmiller et al., 2016). The enzyme encoded by  
21 *Sopen04g001210* did not show any detectable activity against tomato acylsucrose fractions  
22 (Schillmiller et al., 2016), but it may have an acylglucose-specific function.

23 Acylsugar export: It is not known how acylsugars are secreted out of trichome cells, or whether  
24 acylglucose and acylsucrose molecules are exported by the same mechanism. The SCPL GAT,  
25 which catalyzes the second step in phase 2 of acylglucose biosynthesis, has a cleavable N-  
26 terminal 18-aa signal peptide, and acylsugars (or at least acylglucose molecules) have been  
27 suggested to be secreted via a secretory vesicle pathway (Li et al., 1999; Li and Steffens, 2000).  
28 However, to our knowledge, there is no direct evidence to support this hypothesis, and it remains  
29 unknown if the final 2,3,4-tri-*O*-acylglucose molecules are synthesized in secretory vesicles or in  
30 the cytosol. Here, we identified three putative ABC transporters as candidates for acylsugar (or at

1 least acylsucrose) transport based on their strong intramodular connectivity with ASAT and  
2 acylsugar phase 1 genes (Figure 7).

3

#### 4 **Role of central carbon metabolism**

5 NADP-malic enzymes catalyze the oxidative decarboxylation of malate to generate pyruvate  
6 (required for BCAA biosynthesis), along with CO<sub>2</sub> and NADPH, both of which can be used for  
7 SCFA biosynthesis or carbon fixation. Expression profiles of two genes putatively encoding  
8 RUBISCO small subunit and NADP-malic enzyme showed strong positive correlations with  
9 acylsugar accumulation (Supplemental Dataset 4). Putative orthologs of these genes in *S.*  
10 *lycopersicum* show trichome-enriched expression (Ning et al., 2015), consistent with the  
11 suggestion that these enzymes provide some carbon in *S. pennellii* trichomes to support  
12 acylsugar biosynthesis. Strong positive correlations between acylsugar accumulation and  
13 transcript levels of putative invertase (whose putative ortholog is not expressed in isolated *S.*  
14 *lycopersicum* trichomes; Ning et al., 2015) and a monosaccharide transporter is consistent with  
15 the proposal that *S. pennellii* trichomes import at least some photosynthate from underlying leaf  
16 tissues to support acylsugar biosynthesis (Slocombe et al., 2008). Recently, the role of central  
17 carbon metabolism in supporting specialized metabolism in glandular trichomes of *S.*  
18 *lycopersicum* and *S. habrochaites* was investigated, and it was concluded that glandular  
19 trichomes import most of their fixed carbon from underlying leaf tissues, despite being able to  
20 perform some photosynthetic activities (Balcke et al., 2017). A role of glutathione metabolism in  
21 alleviating oxidative stress was also reported. Our results (Table 3; Figure 7 for glutathione S-  
22 transferase) are consistent with these findings.

23

#### 24 **Regulation of acylsugar production**

25 Both comparative transcriptomics and imazapyr treatment studies revealed that transcript levels  
26 for the two phase 2 acylglucose biosynthetic enzymes (UDP-Glc:FA GT and SCPL GAT) were  
27 not positively correlated with acylsugar levels, and any transcriptional regulation of the UDP-  
28 Glc:FA GT is negatively correlated with acylsugar production (Table 2; Table 5). Many

1 biochemical pathways are regulated, at least in part, by a feedback inhibitory mechanism, and  
2 end-product inhibition often acts on an early rate-limiting step at the enzymatic level. However,  
3 it is interesting to note that a negative correlation exists between acylsugar amount and the UDP-  
4 Glc:FA GT transcript level. Overrepresentation of Gene Ontology terms ‘protein kinase activity’  
5 and ‘transcription factor activity’ among genes strongly correlated with the UDP-Glc:FA GT  
6 (Figure 6) suggests that the UDP-Glc:FA GT activity is controlled by both post-translational and  
7 transcriptional regulation. Association of expression patterns of the SCPL GAT with ribosomal  
8 proteins and other components of the translational machinery (Supplemental Figure 8) suggests  
9 that its activity may be regulated at the translational level.

10 On the other hand, expression levels of most genes involved in acyl chain synthesis  
11 (phase 1) and phase 2 acylsucrose metabolism were positively correlated with acylsugar levels  
12 (Figure 2; Supplemental Dataset 4). WGCNA revealed that three genes putatively encoding  
13 AP2-family transcription factors form a strong gene co-expression network with many acylsugar  
14 and flavonoid metabolic genes. Downregulation of these transcription factor genes in response to  
15 imazapyr treatment parallels significant downregulation of most acylsugar and flavonoid  
16 metabolic genes (Table 5; Supplemental Dataset 5). Additionally, putative orthologs of these  
17 transcription factor genes in *S. lycopersicum* exhibit trichome-enriched expression, similar to  
18 many genes involved in acylsugar and flavonoid metabolism (Ning et al., 2015). Together, these  
19 results imply important roles for these transcription factors in specialized metabolic pathways of  
20 *S. pennellii* trichomes.

21 Our results suggest that a majority of the acylsugar metabolic pathway, from acyl chain  
22 synthesis to acylsugar secretion, represents a remarkable co-expressed genetic network, which  
23 may be under a common regulatory mechanism. However, the roles we have proposed for the  
24 candidate genes remain to be confirmed by genetic or biochemical studies. Confirmation of  
25 common regulatory factors, such as the AP2-family transcription factors, will be useful in  
26 genetic manipulation of acylsugar production.

27

## 28 **METHODS**

### 29 **Plant materials**



1 Three biological replicates were used for each accession in the ‘LOW’ vs. ‘HIGH’ comparison,  
2 and for the ‘MEDIUM’ accession LA1302. Four biological replicates were used for each  
3 accession in the LA1920 vs. LA0716 comparison. Seeds of all *S. pennellii* accessions were  
4 obtained from the C. M. Rick Tomato Genetics Resource Center (University of California,  
5 Davis). Seeds were treated with a solution of 20% commercial bleach (1.2% sodium  
6 hypochlorite) for 20 minutes, and then germinated on moistened filter paper. Seedlings were  
7 transferred to soil at the cotyledon stage, and plants were grown in a growth chamber (24°C day/  
8 20°C night temperature, 150 mMol m<sup>-2</sup> s<sup>-1</sup> photosynthetically active radiation, 16 h photoperiod,  
9 and 75% humidity).

10

### 11 **RNA sequencing (RNA-Seq) library preparation and sequencing**

12 Before RNA extraction, secreted acylsugars were removed from the leaf surface of 10-week-old  
13 plants by dipping them in ethanol for 2-3 seconds. Leaves were then immediately frozen in liquid  
14 nitrogen and stored at -80°C until further use. Total RNA was extracted from leaves using the  
15 RNAqueous Total RNA Isolation kit (ThermoFischer). Genomic DNA was removed using the  
16 TURBO DNA-free kit (ThermoFischer). Quality of each RNA sample was analyzed with the  
17 Agilent 2200 TapeStation software A01.04 (Agilent Technologies). RNA-Seq libraries were  
18 prepared from polyA<sup>+</sup>-selected RNA samples using the Illumina TruSeq RNA Library  
19 Preparation kit v2, and then sequenced on the Illumina HiSeq 2500 v4 125x125-bp paired-end  
20 sequencing platform (High Output Mode) following manufacturer’s specifications. Sequencing  
21 was performed at the Texas A&M Genomics and Bioinformatics service center. Sequence cluster  
22 identification, quality prefiltering, base calling and uncertainty assessment were done in real time  
23 using Illumina’s HCS v2.2.68 and RTA v1.18.66.3 software with default parameter setting.  
24 Sequencer.bcl basecall files were demultiplexed and formatted into .fastq files using bcl2fastq  
25 v2.17.1.14 script conFigureBclToFastq.pl. Sequencing reads were submitted to the National  
26 Center for Biotechnology Information (NCBI) Sequence Read Archive under accession number  
27 SRP136022.

28

### 29 **Quality control and mapping of sequencing reads**

1 Phred quality score distributions of sequencing reads were analyzed with FastQC v0.11.4  
2 (Andrews, 2010). All RNA-Seq libraries had average Phred quality scores of more than 35,  
3 indicating greater than 99.97% base call accuracy. To remove all the low-quality, ambiguous ‘N’  
4 nucleotides (Phred quality score= 2, ASCII code= #), the following settings were applied for  
5 quality control of RNA-Seq reads using Trimmomatic v0.35 (Bolger et al., 2014b)-  
6 ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10, SLIDINGWINDOW= 1:14, MINLEN= 80. Using  
7 these settings, more than 98% of the sequencing reads could be retained, and it ensured that none  
8 of the nucleotides in the reads were ambiguous.

9         Processed reads were mapped to the *S. pennellii* genome v2.0 (LA0716) (Bolger et al.,  
10 2014a) using TopHat2 v2.1.0 (Kim et al., 2013). The following parameters were used for the  
11 mapping process: -mate-inner-dist= 0, -mate-std-dev= 50, -read-realign-edit-dist= 0, -read-edit-  
12 dist= 4, -library-type= fr-unstranded, -read-mismatches= 4, -bowtie-n, -min-anchor-len= 8, -  
13 splice-mismatches= 0, -min-intron-length= 50, -max-intron-length= 50000, -max-insertion-  
14 length= 3, -max-deletion-length= 3, -max-multihits= 20, -min-segment-intron= 50, -max-  
15 segment-intron= 50000, -segment-mismatches= 2, -segment-length= 25. A summary of TopHat2  
16 alignment results are given in Supplemental Table 2. Reads mapped to selected loci of interest  
17 were visualized with IGV (Thorvaldsdottir et al., 2013).

18

## 19 **Differential gene expression analyses**

20 Uniquely-mapped reads from TopHat2 were counted for all *S. pennellii* annotated genes using  
21 HTSeq-Count (Anders et al., 2015). The following parameters were used for the counting  
22 process: -f bam, -r name, -s no, -m union, -a 20. These count files were used for identification of  
23 differentially expressed genes using edgeR (Robinson et al., 2010). Genes with very low  
24 expression levels were filtered out before differential testing. For comparison between the  
25 ‘LOW’ and ‘HIGH’ groups, genes with more than 1 count per million (CPM) in at least six  
26 samples were retained for analysis after read counts were adjusted for transcript lengths. Library  
27 sizes were recalculated after the filtering process, and all the samples had post-filter library sizes  
28 of more than 99.70%. Tagwise dispersions were calculated (prior.df= 3), and genes were  
29 identified as differentially expressed when *P* values, corrected for multiple testing using

1 Benjamini-Hochberg multiple testing correction (Benjamini and Hochberg, 1995), were less than  
2 0.05 (false discovery rate < 0.05), and had at least 2-fold expression differences. For comparison  
3 between LA1920 and LA0716, we used the same pipeline, except genes with more than 1 CPM  
4 in at least two samples were retained for analysis, and tagwise dispersions were calculated with  
5 prior.df= 9. In this analysis, all the samples showed more than 99.85% post-filter library sizes.  
6 Lists of differentially expressed genes as determined by edgeR (Robinson et al., 2010) are given  
7 in Supplemental Dataset 1.

8

### 9 **Functional annotation of differentially expressed genes**

10 Transcript sequences of differentially expressed genes were extracted from the annotated gene  
11 models, and Blast2GO (Conesa et al., 2005) was used for the functional annotation of these  
12 transcript sequences. Sequences were searched (BLASTx; word size=3, HSP length cutoff= 33)  
13 against the NCBI nonredundant database (subset Viridiplantae, taxa: 33090) with e-value  $\leq 1.0E-$   
14 10, and top-20 BLAST hits were reported. BLAST hits for each sequence were then mapped  
15 with the Gene Ontology (GO) terms. Next, annotation was performed with the default parameter  
16 settings except the e-value threshold (e-value  $\leq 1.0E-10$ ). This process also generated the  
17 Enzyme Code (EC) numbers from the Kyoto Encyclopedia of Genes and Genomes (KEGG)  
18 pathway. After executing the whole functional annotation process (BLASTx, mapping, and  
19 annotation), transcript sequences were exported with sequence description and GO terms.  
20 Descriptions of transcript sequences were generated based on similarity levels (e-value and  
21 percentage of identity) to the subject genes, as determined by Blast2GO (Conesa et al., 2005).

22

### 23 **Identification of putative orthologs and their trichome-enriched expression**

24 We used reciprocal BLAST to identify putative orthologs of *S. pennellii* genes in *S.*  
25 *lycopersicum*. An all-versus-all BLAST was performed between annotated proteins of *S.*  
26 *pennellii* v2.0 (Bolger et al., 2014a) and *S. lycopersicum* ITAG2.3 (The Tomato Genome  
27 Consortium, 2012; Fernandez-Pozo et al., 2015) with the following parameters- minimum  
28 percentage identity=70, minimum percentage query coverage= 50. Pairs with lowest e-value for

1 each BLASTp comparison were selected as putative orthologs. Gene expression profiles of *S.*  
2 *pennellii* putative orthologs in *S. lycopersicum* were obtained from Ning et al. (2015).

3

#### 4 **Gene co-expression network analysis**

5 Simple correlation analysis among expression profiles (FPKM values in our 29 samples) of  
6 selected genes was performed using the ‘cor’ function in R programming language (R Core  
7 Team, 2014). Gene co-expression network analyses and identification of modules (clusters of  
8 strongly correlated genes) were performed using the R package WGCNA (Langfelder and  
9 Horvath, 2008). Normalized gene expression data (FPKM values) were used as the input, and the  
10 19,378 genes which passed our filtering criteria for minimum expression levels during ‘LOW’  
11 vs. ‘HIGH’ differential testing (*Sopen12g004230* and *Sopen12g004240* were merged) were  
12 selected for WGCNA analysis. A matrix of pairwise Spearman’s rank correlation coefficients  
13 (SRCCs) between these 19,378 genes was created and then transformed into an adjacency  
14 (connection strength) matrix using ‘signed’ network type and a soft threshold ( $\beta$ ) value of 20.  
15 The value of  $\beta$  was determined based on the scale-free topology fit index ( $R^2 > 0.85$ ) and low  
16 mean connectivity. The weighted adjacency matrix was then converted to a topological overlap  
17 matrix (TOM), and genes were hierarchically clustered based on dissimilarity (dissTOM= 1 -  
18 TOM) of the topological overlap. Co-expression modules were defined as the branches of the  
19 clustering tree (dendrogram) using the following parameters in ‘cutreeDynamic’ function:  
20 distM= disTOM, deepSplit=2, minClusterSize= 200. A module eigengene (ME; representative of  
21 gene expression profile in each module) was defined as the first principle component in each  
22 module. Modules with very similar expression profiles as determined by their module  
23 eigengenes were merged (MEDissThres= 0.2), and a final set of 15 modules were obtained. For  
24 each gene, a ‘module membership’ was calculated as the correlation (SRCC) between expression  
25 profile of that gene and module eigengene. Associations between external traits and modules  
26 were measured as determined by SRCC between trait data and module eigengenes; ‘gene  
27 significance’ values for each gene in a module of interest were calculated as the correlation  
28 (SRCC) between expression profile of a gene and external trait data. The top-50 intramodular  
29 connections for selected genes were visualized by VisANT (Hu et al., 2005).

1

## 2 **Inhibitor studies**

3 Compound leaves of *S. pennellii* LA0716 bearing five leaflets were used for the inhibitor study.  
4 Imazapyr (Sigma-Aldrich, USA) was dissolved in distilled water, and administered as 0.1 mM  
5 and 1 mM solutions for 36 hours. Cuttings were transferred to small trays containing imazapyr or  
6 control solutions, and placed under growth chamber conditions. Total RNA was extracted as  
7 described before, and RNA-Seq libraries were prepared using Illumina TruSeq Stranded RNA  
8 LT kit. Libraries were sequenced on the Illumina HiSeq 4000 150x150 bp paired-end sequencing  
9 platform at the Texas A&M Genomics and Bioinformatics service center, College Station. Raw  
10 reads were processed with Trimmomatic (Bolger et al., 2014b), and read-mapping to *S. pennellii*  
11 genome was performed with Tophat2 (Kim et al., 2013) using the same parameters as described  
12 before, except with the following change: -library-type= fr-firststrand. Uniquely-mapped reads  
13 from TopHat2 were counted with HTSeq-Count (Anders et al., 2015) using the following  
14 options: -f bam, -r name, -s reverse, -m union, -a 20. Differentially expressed genes (FDR < 0.05;  
15  $\log_2FC > 1$ ) were identified using edgeR (Robinson et al., 2010) with tagwise dispersions  
16 (prior.df= 12), and genes with more than 1 CPM in at least three samples were retained for the  
17 analysis. Summary results of the inhibitor study are given in Supplemental Dataset 6.

18

## 19 **Accession Numbers**

20 RNA-Seq reads used in this study were submitted to the National Center for Biotechnology  
21 Information (NCBI) Sequence Read Archive under accession numbers SRP136022 and  
22 SRP136034 (imazapyr study).

23

## 24 **Supplemental Data**

25 **Supplemental Figure 1.** Distribution of gene ontology (GO) terms associated with the  
26 1087 common DEGs.

1       **Supplemental Figure 2.** Differential gene expression analyses between low- and high-  
2       acylsugar-producing accessions.

3       **Supplemental Figure 3.** Expression levels of ten selected ‘housekeeping genes’ in  
4       different biological groups.

5       **Supplemental Figure 4.** Biosynthesis of branched-chain and straight-chain acyl  
6       molecules.

7       **Supplemental Figure 5.** Single transcript produced from the locus containing  
8       *Sopen12g004230-Sopen12g004240*.

9       **Supplemental Figure 6.** Correlation among expression profiles of putative defense  
10      genes.

11      **Supplemental Figure 7.** Lys and Arg content of selected proteins encoded by genes co-  
12      expressed with *Sopen11g004040* (FLOWERING LOCUS D; amine oxidase domain).

13      **Supplemental Figure 8.** Full list of enriched gene ontology (GO) terms associated with  
14      genes strongly correlated with the UDP-Glc:FA GT and the SCPL GAT.

15      **Supplemental Figure 9.** Top-50 most strongly connected genes for each of the three  
16      AP2-family transcription factor genes.

17      **Supplemental Figure 10.** Selected *O*-methyltransferase genes on chromosome 6  
18      involved in flavonoid metabolism in *S. pennellii* and *S. lycopersicum*.

19      **Supplemental Table 1.** Amount of acylsugars produced by different accessions of *S.*  
20      *pennellii*.

21      **Supplemental Table 2.** Read mapping (TopHat2 alignment) results with different  
22      accessions of *S. pennellii*.

23      **Supplemental Table 3.** Sequence similarity between Cuphea KAS IV/ KAS II-like  
24      enzymes and *Sopen12g004230-Sopen12g004240*.

1       **Supplemental Table 4.** Selected genes strongly correlated with *Sopen11g004040*  
2       (FLOWERING LOCUS D; amine oxidase domain).

3       **Supplemental Table 5.** Selected genes in WGCNA 'green' module showing strong  
4       correlation with the SCPL GAT.

5       **Supplemental Table 6.** Selected housekeeping genes and translation-related genes  
6       showing correlations in gene expression profiles (GS) with the SCPL GAT.

7       **Supplemental Table 7.** Effect of imazapyr treatment on expression levels of ABC  
8       transporter genes.

9       **Supplemental Table 8.** Effect of imazapyr treatment on expression levels of selected  
10      genes strongly correlated with FLOWERING LOCUS D gene *Sopen11g004040*  
11      (Supplemental Table 4 genes).

12      **Supplemental Dataset 1.** Expression levels of all *S. pennellii* annotated genes and results  
13      of differential gene expression analysis between low- and high-acylsugar-producing  
14      accessions.

15      **Supplemental Dataset 2.** Spearman's rank correlation coefficient (SRCC) among  
16      selected genes' expression profiles in 29 biological samples (different accessions of *S.*  
17      *pennellii*).

18      **Supplemental Dataset 3.** Reciprocal best hits (RBH) between *S. pennellii* and *S.*  
19      *lycopersicum* annotated protein sequences.

20      **Supplemental Dataset 4.** Results of weighted gene correlation network analysis  
21      (WGCNA).

22      **Supplemental Dataset 5.** List of DEGs (between low- and high-acylsugar-producing  
23      accessions) that showed response to imazapyr at both 0.1 mM and 1 mM.

24      **Supplemental Dataset 6.** Summary of imazapyr treatment study.

25

1

2 **Author contributions** – SM designed experiments, prepared samples, analyzed data, and wrote  
3 the manuscript. WJ designed experiments, prepared samples, and reviewed the manuscript. TDM  
4 designed experiments, analyzed data, and wrote the manuscript. All authors read and approved  
5 the final manuscript.

6 **Acknowledgements** – We thank Dr. Charlie Johnson and the staff of the Texas A&M Genomics  
7 and Bioinformatics center for performing Illumina sequencing. We also thank Dr. Rodolfo  
8 Aramayo and Ricardo Perez for technical support on the BioGalaxy server. Most of the sequence  
9 analysis was performed at the Texas A&M University High Performance Research Computer,  
10 where Dr. Michael Dickens provided invaluable assistance. We sincerely appreciate Dr. Alan  
11 Pepper’s assistance with multiple aspects of this project, from initial design to critical review of  
12 the manuscript.

13

14 **Alba, J.M., Montserrat, M., and Fernandez-Munoz, R.** (2009). Resistance to the two-spotted spider mite  
15 (*Tetranychus urticae*) by acylsucroses of wild tomato (*Solanum pimpinellifolium*) trichomes  
16 studied in a recombinant inbred line population. *Exp Appl Acarol* **47**: 35-47.  
17 **Anders, S., Pyl, P.T., and Huber, W.** (2015). HTSeq-a Python framework to work with high-throughput  
18 sequencing data. *Bioinformatics* **31**: 166-169.  
19 **Andrews, S.** (2010). FastQC: a quality control tool for high throughput sequence data.  
20 **Balcke, G., Bennewitz, S., Bergau, N., Athmer, B., Henning, A., Majovsky, P., Jiménez-Gómez, J.M.,**  
21 **Hoehenwarter, W., and Tissier, A.F.** (2017). Multiomics of tomato glandular trichomes reveals  
22 distinct features of central carbon metabolism supporting high productivity of specialized  
23 metabolites. *Plant Cell*.  
24 **Benjamini, Y., and Hochberg, Y.** (1995). Controlling the false discovery rate: a practical and powerful  
25 approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**: 289-300.  
26 **Binder, S.** (2010). Branched-Chain Amino Acid Metabolism in *Arabidopsis thaliana*. *Arabidopsis Book* **8**:  
27 e0137.  
28 **Blauth, S.L., Steffens, J.C., Churchill, G.A., and Mutschler, M.A.** (1999). Identification of QTLs controlling  
29 acylsugar fatty acid composition in an intraspecific population of *Lycopersicon pennellii* (Corr.)  
30 D’Arcy. *Theor. Appl. Genet.* **99**: 373-381.  
31 **Bolger, A., Scossa, F., Bolger, M.E., Lanz, C., Maumus, F., Tohge, T., Quesneville, H., Alseekh, S.,**  
32 **Sorensen, I., Lichtenstein, G., Fich, E.A., Conte, M., Keller, H., Schneeberger, K., Schwacke, R.,**  
33 **Ofner, I., Vrebalov, J., Xu, Y., Osorio, S., Aflitos, S.A., et al.** (2014a). The genome of the stress-  
34 tolerant wild tomato species *Solanum pennellii*. *Nat Genet* **46**: 1034-1038.  
35 **Bolger, A.M., Lohse, M., and Usadel, B.** (2014b). Trimmomatic: a flexible trimmer for Illumina sequence  
36 data. *Bioinformatics* **30**: 2114-2120.  
37 **Bonierbale, M.W., Plaisted, R.L., Pineda, O., and Tanksley, S.D.** (1994). Qtl Analysis of Trichome-  
38 Mediated Insect Resistance in Potato. *Theoretical and Applied Genetics* **87**: 973-987.



- 1 **Brosius, U., Dehmel, T., and Gartner, J.** (2002). Two different targeting signals direct human  
2 peroxisomal membrane protein 22 to peroxisomes. *J. Biol. Chem.* **277**: 774-784.
- 3 **Chortyk, O.T., Kays, S.J., and Teng, Q.** (1997). Characterization of insecticidal sugar esters of *Petunia*. *J.*  
4 *Agric. Food Chem.* **45**: 270-275.
- 5 **Chortyk, O.T., Severson, R.F., Cutler, H.C., and Sisson, V.A.** (1993). Antibiotic Activities of Sugar Esters  
6 Isolated from Selected Nicotiana Species. *Biosci Biotech Bioch* **57**: 1355-1356.
- 7 **Cona, A., Rea, G., Angelini, R., Federico, R., and Tavladoraki, P.** (2006). Functions of amine oxidases in  
8 plant development and defence. *Trends Plant Sci* **11**: 80-88.
- 9 **Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., and Robles, M.** (2005). Blast2GO: a  
10 universal tool for annotation, visualization and analysis in functional genomics research.  
11 *Bioinformatics* **21**: 3674-3676.
- 12 **Consortium, T.T.G.** (2012). The tomato genome sequence provides insights into fleshy fruit evolution.  
13 *Nature* **485**: 635-641.
- 14 **Dehesh, K., Edwards, P., Fillatti, J., Slabaugh, M., and Byrne, J.** (1998). KAS IV: a 3-ketoacyl-ACP  
15 synthase from *Cuphea* sp. is a medium chain specific condensing enzyme. *Plant J* **15**: 383-390.
- 16 **Fan, P., Miller, A.M., Liu, X., Jones, A.D., and Last, R.L.** (2017). Evolution of a flipped pathway creates  
17 metabolic innovation in tomato trichomes through BAHD enzyme promiscuity. *Nat. Commun.* **8**:  
18 2080.
- 19 **Fan, P., Miller, A.M., Schillmiller, A.L., Liu, X., Ofner, I., Jones, A.D., Zamir, D., and Last, R.L.** (2016). In  
20 vitro reconstruction and analysis of evolutionary variation of the tomato acylsucrose metabolic  
21 network. *Proc Natl Acad Sci U S A* **113**: E239-248.
- 22 **Fernandez-Pozo, N., Menda, N., Edwards, J.D., Saha, S., Tecle, I.Y., Strickler, S.R., Bombarely, A.,  
23 Fisher-York, T., Pujar, A., Foerster, H., Yan, A., and Mueller, L.A.** (2015). The Sol Genomics  
24 Network (SGN)-from genotype to phenotype to breeding. *Nucleic Acids Res.* **43**: D1036-1041.
- 25 **Figueiredo, A., Monteiro, F., and Sebastiana, M.** (2014). Subtilisin-like proteases in plant-pathogen  
26 recognition and immune priming: a perspective. *Front Plant Sci* **5**: 739.
- 27 **Fobes, J.F., Mudd, J.B., and Marsden, M.P.** (1985). Epicuticular Lipid Accumulation on the Leaves of  
28 *Lycopersicon pennellii* (Corr.) D'Arcy and *Lycopersicon esculentum* Mill. *Plant Physiol* **77**: 567-  
29 570.
- 30 **Ghangas, G.S., and Steffens, J.C.** (1993). UDPglucose: fatty acid transglucosylation and transacylation in  
31 triacylglycerol biosynthesis. *Proc Natl Acad Sci U S A* **90**: 9911-9915.
- 32 **Halitschke, R., and Baldwin, I.T.** (2003). Antisense LOX expression increases herbivore performance by  
33 decreasing defense responses and inhibiting growth-related transcriptional reorganization in  
34 *Nicotiana attenuata*. *Plant J.* **36**: 794-807.
- 35 **Hare, J.D.** (2005). Biological activity of acyl glucose esters from *Datura wrightii* glandular trichomes  
36 against three native insect herbivores. *J. Chem. Ecol.* **31**: 1475-1491.
- 37 **Hawthorne, D.J., Shapiro, J.A., Tingey, W.M., and Mutschler, M.A.** (1992). Trichome-Borne and  
38 Artificially Applied Acylsugars of Wild Tomato Deter Feeding and Oviposition of the Leafminer  
39 *Liriomyza-Trifolii*. *Entomol Exp Appl* **65**: 65-73.
- 40 **Hill, K., and Rhode, O.** (1999). Sugar-based surfactants for consumer products and technical  
41 applications. *Fett-Lipid* **101**: 25-33.
- 42 **Hu, Z., Mellor, J., Wu, J., Yamada, T., Holloway, D., and Delisi, C.** (2005). VisANT: data-integrating visual  
43 framework for biological networks and modules. *Nucleic Acids Res.* **33**: W352-357.
- 44 **Jackson, P.A., Galinha, C.I., Pereira, C.S., Fortunato, A., Soares, N.C., Amancio, S.B., and Pinto Ricardo,  
45 C.P.** (2001). Rapid deposition of extensin during the elicitation of grapevine callus cultures is  
46 specifically catalyzed by a 40-kilodalton peroxidase. *Plant Physiol* **127**: 1065-1076.

- 1 **Juvik, J.A., Shapiro, J.A., Young, T.E., and Mutschler, M.A.** (1994). Acylglucoses from Wild Tomatoes  
2 Alter Behavior and Reduce Growth and Survival of *Helicoverpa-Zea* and *Spodoptera-Exigua*  
3 (*Lepidoptera, Noctuidae*). *Journal of Economic Entomology* **87**: 482-492.
- 4 **Kim, D., Perte, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L.** (2013). TopHat2: accurate  
5 alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome*  
6 *Biol* **14**.
- 7 **Kim, J., Matsuba, Y., Ning, J., Schillmiller, A.L., Hammar, D., Jones, A.D., Pichersky, E., and Last, R.L.**  
8 (2014). Analysis of natural and induced variation in tomato glandular trichome flavonoids  
9 identifies a gene not present in the reference genome. *Plant Cell* **26**: 3272-3285.
- 10 **Kroumova, A.B., Zaitlin, D., and Wagner, G.J.** (2016). Natural variability in acyl moieties of sugar esters  
11 produced by certain tobacco and other *Solanaceae* species. *Phytochemistry* **130**: 218-227.
- 12 **Kuai, J.P., Ghangas, G.S., and Steffens, J.C.** (1997). Regulation of Triacylglyceride Fatty Acid Composition  
13 (Uridine Diphosphate Glucose:Fatty Acid Glucosyltransferases with Overlapping Chain-Length  
14 Specificity). *Plant Physiol* **115**: 1581-1587.
- 15 **La Camera, S., Gouzerh, G., Dhondt, S., Hoffmann, L., Fritig, B., Legrand, M., and Heitz, T.** (2004).  
16 Metabolic reprogramming in plant innate immunity: the contributions of phenylpropanoid and  
17 oxylipin pathways. *Immunol Rev* **198**: 267-284.
- 18 **Langfelder, P., and Horvath, S.** (2008). WGCNA: an R package for weighted correlation network analysis.  
19 *BMC Bioinformatics* **9**: 559.
- 20 **Lawson, D.M., Lunde, C.F., and Mutschler, M.A.** (1997). Marker-assisted transfer of acylsugar-mediated  
21 pest resistance from the wild tomato, *Lycopersicon pennellii*, to the cultivated tomato,  
22 *Lycopersicon esculentum*. *Mol Breeding* **3**: 307-317.
- 23 **Leckie, B.M., De Jong, D.M., and Mutschler, M.A.** (2012). Quantitative trait loci increasing acylsugars in  
24 tomato breeding lines and their impacts on silverleaf whiteflies. *Mol Breeding* **30**: 1621-1634.
- 25 **Li, A.X., and Steffens, J.C.** (2000). An acyltransferase catalyzing the formation of diacylglyceride is a serine  
26 carboxypeptidase-like protein. *Proc Natl Acad Sci U S A* **97**: 6902-6907.
- 27 **Li, A.X., Eannetta, N., Ghangas, G.S., and Steffens, J.C.** (1999). Glucose polyester biosynthesis.  
28 Purification and characterization of a glucose acyltransferase. *Plant Physiol* **121**: 453-460.
- 29 **Liedl, B.E., Lawson, D.M., White, K.K., Shapiro, J.A., Cohen, D.E., Carson, W.G., Trumble, J.T., and**  
30 **Mutschler, M.A.** (1995). Acylsugars of Wild Tomato *Lycopersicon-Pennellii* Alters Settling and  
31 Reduces Oviposition of *Bemisia-Argentifolii* (*Homoptera, Aleyrodidae*). *Journal of Economic*  
32 *Entomology* **88**: 742-748.
- 33 **Luu, V.T., Weinhold, A., Ullah, C., Dressel, S., Schoettner, M., Gase, K., Gaquerel, E., Xu, S., and**  
34 **Baldwin, I.T.** (2017). O-acyl sugars protect a wild tobacco from both native fungal pathogens  
35 and a specialist herbivore. *Plant Physiol.* **174**: 370-386.
- 36 **Maloney, G.S., Kochevenko, A., Tieman, D.M., Tohge, T., Krieger, U., Zamir, D., Taylor, M.G., Fernie,**  
37 **A.R., and Klee, H.J.** (2010). Characterization of the branched-chain amino acid aminotransferase  
38 enzyme family in tomato. *Plant Physiol* **153**: 925-936.
- 39 **Manabe, Y., Tinker, N., Colville, A., and Miki, B.** (2007). CSR1, the sole target of imidazolinone herbicide  
40 in *Arabidopsis thaliana*. *Plant Cell Physiol* **48**: 1340-1358.
- 41 **Marcos, J.F., Munoz, A., Perez-Paya, E., Misra, S., and Lopez-Garcia, B.** (2008). Identification and  
42 rational design of novel antimicrobial peptides for plant protection. *Annu Rev Phytopathol* **46**:  
43 273-301.
- 44 **McDowell, E.T., Kapteyn, J., Schmidt, A., Li, C., Kang, J.H., Descour, A., Shi, F., Larson, M., Schillmiller,**  
45 **A., An, L., Jones, A.D., Pichersky, E., Soderlund, C.A., and Gang, D.R.** (2011). Comparative  
46 functional genomic analysis of *Solanum* glandular trichome types. *Plant Physiol.* **155**: 524-539.

- 1 **Moghe, G.D., Leong, B.J., Hurney, S.M., Daniel Jones, A., and Last, R.L.** (2017). Evolutionary routes to  
2 biochemical innovation revealed by integrative analysis of a plant-defense related specialized  
3 metabolic pathway. *eLife* **6**.
- 4 **Mueller, S., Hilbert, B., Dueckershoff, K., Roitsch, T., Krischke, M., Mueller, M.J., and Berger, S.** (2008).  
5 General detoxification and stress responses are mediated by oxidized lipids through TGA  
6 transcription factors in Arabidopsis. *Plant Cell* **20**: 768-785.
- 7 **Niebel, A., De Almeida Engler, J., Tire, C., Engler, G., Van Montagu, M., and Gheysen, G.** (1993).  
8 Induction Patterns of an Extensin Gene in Tobacco upon Nematode Infection. *Plant Cell* **5**: 1697-  
9 1710.
- 10 **Ning, J., Moghe, G.D., Leong, B., Kim, J., Ofner, I., Wang, Z., Adams, C., Jones, A.D., Zamir, D., and Last,**  
11 **R.L.** (2015). A Feedback-Insensitive Isopropylmalate Synthase Affects Acylsugar Composition in  
12 Cultivated and Wild Tomato. *Plant Physiol* **169**: 1821-1835.
- 13 **Noguchi, A., Sasaki, N., Nakao, M., Fukami, H., Takahashi, S., Nishino, T., and Nakayama, T.** (2008).  
14 cDNA cloning of glycosyltransferases from Chinese wolfberry (*Lycium barbarum* L.) fruits and  
15 enzymatic synthesis of a catechin glucoside using a recombinant enzyme (UGT73A10). *J. Mol.*  
16 *Catal. B Enzym.* **55**: 84-92.
- 17 **Perez-Castorena, A.L., Martinez, M., and Maldonado, E.** (2010). Labdanes and sucrose esters from  
18 *Physalis sordida*. *J Nat Prod* **73**: 1271-1276.
- 19 **Prost, I., Dhondt, S., Rothe, G., Vicente, J., Rodriguez, M.J., Kift, N., Carbonne, F., Griffiths, G.,**  
20 **Esquerre-Tugaye, M.T., Rosahl, S., Castresana, C., Hamberg, M., and Fournier, J.** (2005).  
21 Evaluation of the antimicrobial activities of plant oxylipins supports their involvement in defense  
22 against pathogens. *Plant Physiol* **139**: 1902-1913.
- 23 **Puterka, G.J., Farone, W., Palmer, T., and Barrington, A.** (2003). Structure-function relationships  
24 affecting the insecticidal and miticidal activity of sugar esters. *Journal of Economic Entomology*  
25 **96**: 636-644.
- 26 **Reumann, S., Quan, S., Aung, K., Yang, P., Manandhar-Shrestha, K., Holbrook, D., Linka, N.,**  
27 **Switzenberg, R., Wilkerson, C.G., Weber, A.P., Olsen, L.J., and Hu, J.** (2009). In-depth proteome  
28 analysis of Arabidopsis leaf peroxisomes combined with in vivo subcellular targeting verification  
29 indicates novel metabolic and regulatory functions of peroxisomes. *Plant Physiol* **150**: 125-143.
- 30 **Robinson, M.D., McCarthy, D.J., and Smyth, G.K.** (2010). edgeR: a Bioconductor package for differential  
31 expression analysis of digital gene expression data. *Bioinformatics* **26**: 139-140.
- 32 **Schillmiller, A.L., Charbonneau, A.L., and Last, R.L.** (2012). Identification of a BAHG acetyltransferase  
33 that produces protective acyl sugars in tomato trichomes. *Proc Natl Acad Sci U S A* **109**: 16377-  
34 16382.
- 35 **Schillmiller, A.L., Gilgallon, K., Ghosh, B., Jones, A.D., and Last, R.L.** (2016). Acylsugar Acylhydrolases:  
36 Carboxylesterase-Catalyzed Hydrolysis of Acylsugars in Tomato Trichomes. *Plant Physiol* **170**:  
37 1331-1344.
- 38 **Schillmiller, A.L., Moghe, G.D., Fan, P., Ghosh, B., Ning, J., Jones, A.D., and Last, R.L.** (2015). Functionally  
39 divergent alleles and duplicated Loci encoding an acyltransferase contribute to acylsugar  
40 metabolite diversity in *Solanum* trichomes. *Plant Cell* **27**: 1002-1017.
- 41 **Schmidt, A., Li, C., Jones, A.D., and Pichersky, E.** (2012). Characterization of a flavonol 3-O-  
42 methyltransferase in the trichomes of the wild tomato species *Solanum habrochaites*. *Planta*  
43 **236**: 839-849.
- 44 **Schutt, B.S., Abbadì, A., Loddenkotter, B., Brummel, M., and Spener, F.** (2002). Beta-ketoacyl-acyl  
45 carrier protein synthase IV: a key enzyme for regulation of medium-chain fatty acid synthesis in  
46 *Cuphea lanceolata* seeds. *Planta* **215**: 847-854.
- 47 **Shapiro, J.A., Steffens, J.C., and Mutschler, M.A.** (1994). Acylsugars of the Wild Tomato *Lycopersicon*  
48 *Pennellii* in Relation to Geographic-Distribution of the Species. *Biochem Syst Ecol* **22**: 545-561.

- 1 **Shockey, J.M., Fulda, M.S., and Browse, J.** (2003). Arabidopsis contains a large superfamily of acyl-  
2 activating enzymes. Phylogenetic and biochemical analysis reveals a new class of acyl-coenzyme  
3 a synthetases. *Plant Physiol* **132**: 1065-1076.
- 4 **Singh, V., Roy, S., Giri, M.K., Chaturvedi, R., Chowdhury, Z., Shah, J., and Nandi, A.K.** (2013).  
5 *Arabidopsis thaliana* FLOWERING LOCUS D is required for systemic acquired resistance. *Mol.*  
6 *Plant Microbe Interact.* **26**: 1079-1088.
- 7 **Slabaugh, M.B., Leonard, J.M., and Knapp, S.J.** (1998). Condensing enzymes from *Cuphea wrightii*  
8 associated with medium chain fatty acid biosynthesis. *Plant J* **13**: 611-620.
- 9 **Slocombe, S.P., Schauvinhold, I., McQuinn, R.P., Besser, K., Welsby, N.A., Harper, A., Aziz, N., Li, Y.,**  
10 **Larson, T.R., Giovannoni, J., Dixon, R.A., and Broun, P.** (2008). Transcriptomic and reverse  
11 genetic analyses of branched-chain fatty acid and acyl sugar production in *Solanum pennellii* and  
12 *Nicotiana benthamiana*. *Plant Physiol* **148**: 1830-1846.
- 13 **Team, R.C.** (2014). R Core Team. R: A language and environment for statistical computing. Vienna,  
14 Austria: R Foundation for Statistical Computing.
- 15 **Thorvaldsdottir, H., Robinson, J.T., and Mesirov, J.P.** (2013). Integrative Genomics Viewer (IGV): high-  
16 performance genomics data visualization and exploration. *Brief Bioinform* **14**: 178-192.
- 17 **Walters, D.S., and Steffens, J.C.** (1990). Branched Chain Amino Acid Metabolism in the Biosynthesis of  
18 *Lycopersicon pennellii* Glucose Esters. *Plant Physiol* **93**: 1544-1551.
- 19 **Yamada, K., Osakabe, Y., Mizoi, J., Nakashima, K., Fujita, Y., Shinozaki, K., and Yamaguchi-Shinozaki, K.**  
20 (2010). Functional analysis of an *Arabidopsis thaliana* abiotic stress-inducible facilitated diffusion  
21 transporter for monosaccharides. *J Biol Chem* **285**: 1138-1146.

22

23

## 24 **FIGURE LEGENDS**

25

26 **Figure 1.** Current model of acylsugar production in *Solanum pennellii* and workflow to identify  
27 candidate genes for acylsugar metabolism. **(A)** Branched-chain fatty acids (BCFAs) are derived  
28 from branched-chain amino acids (BCAAs), and straight-chain fatty acids (SCFAs) are assumed  
29 to be produced by fatty acid synthase (FAS). UDP-glucose:fatty acid glucosyltransferase (UDP-  
30 Glc:FA GT) uses free fatty acids to initiate acylglucose biosynthesis, whereas acylsugar  
31 acyltransferase 1 (ASAT1) uses fatty acyl-CoAs to initiate acylsucrose biosynthesis. Enzymes  
32 are highlighted in red. Unidentified enzymes and transporters are marked with “?”. Double  
33 arrows indicate more than one enzymatic step. C1-4 indicates esterification at respective  
34 positions on glucose. P2-4 indicates esterification at respective positions on the sucrose pyranose  
35 ring. **(B)** Workflow to identify acylsugar candidate genes. Comparative transcriptomics  
36 identified 1679 differentially expressed genes (DEGs) between three low-acylsugar-producing

1 accessions ('LOW') and three high-acylsugar-producing accessions ('HIGH'). An independent  
2 comparison with higher sequencing coverage identified 3524 DEGs between another low-  
3 acylsugar-producing accession, LA1920, and the high-acylsugar-producing accession LA0716.  
4 1087 DEGs were common to both comparisons. Expression levels of selected candidate genes  
5 were determined in LA1302, which accumulates an intermediate level of acylsugars  
6 ('MEDIUM'). Weighted gene correlation network analysis (WGCNA) (Langfelder and Horvath,  
7 2008) was performed to identify co-expressed gene networks associated with acylsugar  
8 accumulation. Finally, analysis of gene expression in response to inhibition of BCAA  
9 biosynthesis narrowed the list of candidates.

10

11 **Figure 2.** Heatmaps showing expression levels of genes with known and putative functions in  
12 acylsugar metabolism. Genes are designated by their gene identifier numbers (Sopen IDs). (A)  
13 Acylsugar phase 1-related genes (Table 1). Expression levels of many candidate genes in  
14 LA1302 ('MEDIUM' accession) were intermediate between low- and high-acylsugar-producing  
15 accessions. (B) Acylsugar phase 2-related genes (Table 2). *Sopen04g001210* (marked with an  
16 asterisk) was identified as a carboxylesterase gene related to ASH, but the encoded protein does  
17 not exhibit in vitro activity against tomato acylsucrose fractions (Schillmiller et al., 2016). (C)  
18 Other genes related to acylsugar metabolism (Table 3). Two genes (*Sopen04g023150* and  
19 *Sopen01g047950*) were removed from the heatmap due to their very high expression levels.  
20 *Sopen04g023150* had FPKM values of 570, 624, 455, 179, and 254 in 'LOW', LA1920,  
21 LA1302, LA0716, and 'HIGH' groups, respectively. *Sopen01g047950* had FPKM values of 61,  
22 203, 79, 29, and 19 in 'LOW', LA1920, LA1302, LA0716, and 'HIGH' groups,  
23 respectively. Abbreviations: FPKM= fragments per kilobase of transcript per million mapped  
24 reads; BCAA= branched-chain amino acid; FAS= fatty acid synthase components; AAE1= acyl-  
25 activating enzyme 1; AG= acylglucose phase 2; ASAT= acylsucrose acyltransferase; ASH=  
26 acylsugar hydrolase; ABC= ABC transporters; CCM= central carbon metabolism.

27

28 **Figure 3.** Correlation among expression profiles of selected genes. FPKM values for 28 phase 1-  
29 and phase 2-related genes (from Table 1 and Table 2, respectively) in our 29 samples were used

1 to determine pairwise Spearman's rank correlation coefficients (SRCC). A [28x28] matrix was  
2 created to visualize correlation. SRCC values and associated P values are given in Supplemental  
3 Dataset 2. *Sopen04g001210* (marked with an asterisk) was identified as a carboxylesterase gene  
4 related to ASH (Schillmiller et al., 2016). Abbreviations: BCAA= branched-chain amino acid;  
5 FAS= fatty acid synthase components; AAE1= acyl-activating enzyme 1; AG= acylglucose  
6 phase 2; ASAT= acylsucrose acyltransferase; ASH= acylsugar hydrolase.

7

8 **Figure 4.** Heatmap showing expression levels of genes putatively involved in plant defense  
9 (Table 4 genes). Expression levels of many genes in LA1302 were similar to low-acylsugar-  
10 producing accessions but different from high-acylsugar-producing accessions. Some oxylipin-  
11 related defense genes, which were upregulated in low-acylsugar-producing accessions, had  
12 higher expression levels than amine oxidase-related defense genes, which were upregulated in  
13 high-acylsugar-producing accessions. Abbreviations: FPKM= fragments per kilobase of  
14 transcript per million mapped reads; FAD= fatty acid desaturase; LOX= lipoxygenase; GST=  
15 glutathione S-transferase; OLM= oxylipin metabolism; SLP= subtilisin-like protease; AOX=  
16 amine oxidase.

17

18 **Figure 5.** Weighted gene correlation network analysis (WGCNA) (Langfelder and Horvath,  
19 2008). (A) Dendrogram clustering 19,378 genes into 15 co-expressed modules based on their  
20 expression profiles in all 29 biological samples (see Methods section). For each module, a  
21 module eigengene was selected by WGCNA as representative of the expression profile for that  
22 module. (B) Association of module eigengenes (MEs) with acylsugar level and seven selected  
23 gene expression profiles. The seven genes encode acylglucose biosynthetic enzymes (UDP-  
24 Glc:FA GT and SCPL GAT), acylsucrose biosynthetic enzymes (ASAT1, ASAT2, and ASAT3),  
25 and defense related proteins (LOX and FLD as representatives of oxylipin-related and amine  
26 oxidase-related defense networks, respectively). Spearman's rank correlation coefficient (SRCC)  
27 values between MEs and traits are shown; P values associated with SRCC are given in  
28 parentheses. Abbreviations: UDP-Glc:FA GT= UDP-glucose:fatty acid glucosyltransferase;

1 SCPL GAT= serine carboxypeptidase-like glucose acyltransferase; ASAT= acylsugar  
2 acyltransferase; LOX= lipoxygenase; FLD= FLOWERING LOCUS D.

3

4 **Figure 6.** Enrichment analysis of Gene Ontology (GO) terms associated with genes most  
5 strongly correlated with two genes encoding known acylglucose biosynthetic enzymes (UDP-  
6 Glc:FA GT and SCPL GAT). Enrichment analysis was performed with Blast2GO (Conesa et al.,  
7 2005) using Fischer's exact test ( $P < 0.05$ ). Top-100 genes associated with *Sopen01g049990*  
8 (UDP-Glc:FA GT) in both 'lightcyan' and 'tan' modules, and top-100 genes associated with  
9 *Sopen10g020280* (SCPL GAT) in 'green' module were used for the analysis. Categories are  
10 listed from top to bottom based on their P values (from  $P = 3.57E-06$  for 'RNA binding' to  $P =$   
11  $4.39E-02$  for 'GTP binding'). GO terms in 'molecular function' subsection only were selected,  
12 and redundant categories were combined. A full list of GO terms and associated P values are  
13 given in Supplemental Figure 8.

14

15 **Figure 7.** Simplified acylsugar metabolic gene network. Nodes and edges represent genes and  
16 intramodular connectivities, respectively. We used WGCNA to identify the top-50 genes most  
17 strongly connected to each of the eight transcription factor genes (all green squares except  
18 *Sopen03g030650* and *Sopen05g031840*) that responded to imazapyr and were in the 'turquoise'  
19 module, as well as three acylsucrose biosynthetic genes (three ASATs). These 11 top-50 lists  
20 contained only 160 different genes. Using VisANT (Hu et al., 2005), with an edge weight cutoff  
21 value of 0.20, 41 genes were identified as most strongly connected. Three genes putatively  
22 encoding AP2-family transcription factors (*Sopen05g008450*, *Sopen12g021250*, and  
23 *Sopen03g036630*) showed strongest connectivity with acylsugar (and also flavonoid) metabolic  
24 genes. This network also includes two additional transcription factor genes (*Sopen03g030650*  
25 and *Sopen05g031840*). *Sopen03g030650* was downregulated only at 1 mM imazapyr treatment,  
26 whereas *Sopen05g031840* was slightly upregulated at both 0.1 and 1 mM imazapyr treatment.

27

1 **Figure 8.** Proposed model of acylsugar metabolism in *S. pennellii*. Role of different subcellular  
 2 compartments are shown. Metabolites are highlighted in blue within boxes. Enzymes,  
 3 transporters and transcription factors (TFs) with known and putative functions in acylsugar  
 4 metabolism are highlighted in black within ovals. Single arrow does not necessarily represent a  
 5 single enzymatic step. This figure was prepared using WikiPathways (Pico et al., 2008).  
 6 Abbreviations: NADP:ME= NADP-malic enzyme; 3-PG= 3-phosphoglycerate; BCKA=  
 7 branched-chain keto acid.

8

9 **TABLES**

10 **Table 1.** Differentially expressed genes involved in phase 1 of acylsugar production (acyl chain  
 11 synthesis).

Gene ID	'LOW' vs. 'HIGH'		LA1920 vs. LA0716		Annotation
	Log <sub>2</sub> FC	FDR	Log <sub>2</sub> FC	FDR	
Branched-chain amino acid (BCAA)/ Branched-chain fatty acid (BCFA) metabolism					
Sopen11g004560	1.44	4.30E-05	1.63	2.20E-05	Acetolactate synthase small subunit
Sopen07g027240	1.46	1.00E-11	1.38	2.80E-06	Ketol-acid reductoisomerase
Sopen05g032060	1.78	1.80E-10	1.71	4.50E-06	Dihydroxy-acid dehydratase
Sopen08g005060	1.01	3.10E-02	1.38	3.60E-07	Isopropylmalate synthase
Sopen08g005140	4.97	3.30E-13	5.12	6.40E-14	Isopropylmalate synthase
Sopen04g030820	5.28	3.50E-13	4.55	4.40E-09	Branched-chain aminotransferase-2
Sopen04g026270	3.37	2.80E-12	2.66	2.10E-07	Branched-chain keto acid dehydrogenase E1 subunit
Sopen01g028100	1.99	1.40E-16	1.53	9.80E-08	Branched-chain keto acid dehydrogenase E2 subunit
Sopen07g023250	4.75	1.40E-08	3.87	2.30E-04	3-Hydroxyisobutyryl-CoA hydrolase
Sopen05g023470	1.34	1.40E-02	1.08	1.00E-02	3-Hydroxyisobutyryl-CoA hydrolase
Sopen12g032690	2.02	7.80E-03	1.87	1.30E-02	Mitochondrial acyl-CoA thioesterase
Fatty acid synthase (FAS) components; * misannotated as two separate genes					
Sopen12g004230 *	5.28	4.40E-18	5.19	1.40E-12	Beta-ketoacyl-ACP synthase II (N-terminus)



Sopen12g004240 *	4.73	1.50E-26	4.39	8.90E-14	Reverse transcriptase; KAS IV/ KAS II-like domain (C-terminus)
Sopen08g002520	4.56	1.40E-13	3.67	3.70E-07	Beta-ketoacyl-ACP synthase III
Sopen05g009610	3.32	3.80E-11	2.73	1.80E-08	Beta-ketoacyl-ACP reductase
Sopen12g029240	3.34	1.60E-09	2.85	2.90E-08	Enoyl-ACP reductase domain
Acyl-activating enzymes					
Sopen02g027670	3.08	5.50E-10	2.49	2.70E-05	Acyl-activating enzyme 1
Sopen02g027680	5.29	5.90E-12	4.23	3.70E-06	Acyl-activating enzyme 1
Sopen07g023200	4.16	2.80E-07	3.72	5.30E-05	Acyl-activating enzyme 1
Sopen07g023220	5.81	3.80E-27	4.02	3.60E-06	Acyl-activating enzyme 1

1 Log<sub>2</sub>FC and FDR indicate log<sub>2</sub> (fold-change) and false discovery rate (*P* values adjusted for  
 2 multiple-testing), respectively. Positive and negative log<sub>2</sub>FC values indicate higher and lower  
 3 expression levels, respectively in high-acylsugar-producing accessions.

4

5 **Table 2.** Genes involved in phase 2 of acylsugar production.

Gene ID	'LOW' vs. 'HIGH'		LA1920 vs. LA0716		Annotation
	Log <sub>2</sub> FC	FDR	Log <sub>2</sub> FC	FDR	
Acylglucose biosynthesis					
Sopen01g049990	-1.05	2.70E-03	-0.01	0.99	UDP-glucose:fatty acid glucosyltransferase
Sopen10g020280	-0.01	0.99	-0.01	0.99	Serine carboxypeptidase-like glucose acyltransferase
Acylsugar acyltransferase (acylsucrose biosynthesis)					
Sopen12g002290	5.11	4.10E-14	4.47	1.10E-06	Acylsugar acyltransferase 1
Sopen04g006140	4.5	9.60E-06	4.53	4.60E-06	Acylsugar acyltransferase 2
Sopen11g026960	4.95	3.20E-08	4.28	3.20E-05	Acylsugar acyltransferase 3
Acylsugar hydrolase (acylsucrose metabolism) and the related carboxylesterase					
Sopen05g030120	-0.44	0.37	1.04	1.8E-03	Acylsugar acylhydrolase 1
Sopen05g030130	3.84	1.70E-08	4.02	1.30E-06	Acylsugar acylhydrolase 2
Sopen09g030520	3.33	8.80E-15	3.12	8.50E-05	Acylsugar acylhydrolase 3
Sopen04g001210	4.69	9.50E-08	4.6	1.80E-05	Carboxylesterase (alpha/beta

					hydrolase fold)
--	--	--	--	--	-----------------

1 Log<sub>2</sub>FC and FDR indicate log<sub>2</sub> (fold-change) and false discovery rate (*P* values adjusted for  
 2 multiple-testing), respectively. Positive and negative log<sub>2</sub>FC values indicate higher and lower  
 3 expression levels, respectively in high-acylsugar-producing accessions.

4

5 **Table 3.** Differentially expressed genes involved in transport process, central carbon  
 6 metabolism, and regulation of gene expression.

Gene ID	'LOW' vs. 'HIGH'		LA1920 vs. LA0716		Annotation
	Log <sub>2</sub> F C	FDR	Log <sub>2</sub> F C	FDR	
Transporters					
Sopen04g023150	-1.16	2.00E-02	-1.82	4.40E-09	ABC transporter F family member
Sopen01g047950	-1.68	2.10E-02	-2.82	3.30E-04	ABC transporter G family member
Sopen04g005380	3.19	2.80E-08	2.8	6.50E-04	ABC transporter G family member
Sopen12g034820	4.37	2.10E-12	3.47	4.00E-06	Pleiotropic drug resistance protein 1- like (ABC transporter G family)
Sopen03g001870	4.52	1.10E-32	4	3.60E-09	ABC transporter B family member
Central carbon metabolism					
Sopen07g006810	5.01	6.70E-12	3.63	4.30E-04	RUBISCO small subunit
Sopen08g020180	2.11	1.90E-10	2.18	2.30E-07	NADP-dependent malic enzyme
Sopen03g040490	2.99	9.60E-13	2.87	1.10E-11	Invertase (beta-fructofuranosidase)
Sopen01g042050	3.39	1.30E-09	2.76	3.10E-06	Sugar transporter ERD6-like
Selected transcription factors *					
Sopen05g008450	4.9	1.3E-21	4.15	6.3E-10	AP2 domain
Sopen12g021250	3.9	4.4E-15	2.88	5.6E-06	AP2 domain
Sopen03g036630	3.13	1.5E-12	2.71	3.8E-07	AP2 domain
Sopen10g031080	2.68	5.5E-12	2.53	3E-07	Homeobox associated leucine zipper
Sopen06g024660	2.88	1.9E-09	4.02	9.8E-11	Myb-like DNA-binding domain
Sopen08g028640	3.54	1.5E-07	3.57	3.6E-08	AP2 domain
Sopen02g021600	1.79	5.6E-05	2.07	6.4E-05	Myb-like DNA-binding domain
Sopen01g037680	1.8	9E-04	1.39	2.1E-02	TCP family transcription factor

Sopen03g037210	1.65	1.1E-03	2.62	9.7E-05	Helix-loop-helix DNA-binding domain
----------------	------	---------	------	---------	-------------------------------------

1 Log<sub>2</sub>FC and FDR indicate log<sub>2</sub> (fold-change) and false discovery rate (*P* values adjusted for  
2 multiple-testing), respectively. Positive and negative log<sub>2</sub>FC values indicate higher and lower  
3 expression levels, respectively in high-acylsugar-producing accessions. Transcription factor  
4 genes (marked with an asterisk) were selected based on their response to imazapyr treatment (see  
5 Table 5).

6

7 **Table 4.** Differentially expressed genes putatively involved in oxylipin-related and amine  
8 oxidase-related defense systems.

Gene ID	'LOW' vs. 'HIGH'		LA1920 vs. LA0716		Annotation
	Log <sub>2</sub> FC	FDR	Log <sub>2</sub> FC	FDR	
Desaturase and lipoxygenase					
Sopen07g032620	-2.43	2.10E-08	-3.12	4.10E-09	Fatty acid desaturase 4
Sopen11g004460	-1.37	2.30E-03	-3.86	1.40E-30	Stearoyl-(ACP)-9-desaturase
Sopen04g016350	-4.23	7.00E-22	-3.00	2.30E-05	Delta(12)-fatty acid desaturase
Sopen00g008480	-6.29	9.10E-08	-8.72	4.60E-67	Delta(12)-fatty acid desaturase
Sopen12g023440	-2.71	5.60E-05	-2.03	1.60E-02	Delta(12)-fatty acid desaturase
Sopen12g034870	-2.69	4.10E-04	-2.48	1.00E-02	Delta(12)-fatty acid desaturase
Sopen01g002520	-7.99	2.40E-15	-10.7	2.30E-40	Linoleate 13S-lipoxygenase
Sopen09g030670	-1.38	3.30E-08	-1.84	1.90E-10	Linoleate 9S-lipoxygenase
Sopen00g005130	2.03	9.70E-09	1.97	1.00E-05	Linoleate 9S-lipoxygenase
Oxylipin signaling and detoxification					
Sopen02g021560	-3.55	6.80E-10	-4.35	3.40E-10	Transcription factor TGA5-like
Sopen10g035600	-1.06	5.10E-05	-1.63	5.50E-11	12-Oxo-phytodienoate reductase
Sopen09g006370	-7.11	2.20E-16	-11.1	1.40E-08	Glutathione S-transferase-like
Sopen10g017090	-4.89	7.00E-09	-7.08	7.10E-15	Glutathione S-transferase
Sopen09g006360	-1.83	2.10E-03	-1.42	3.30E-02	Glutathione S-transferase
Sopen12g026980*	2.72	1.80E-09	1.21	3.40E-04	Glutathione S-transferase
Other plant defense-related					
Sopen08g028270	-7.03	2.90E-16	-4.15	8.60E-05	Subtilisin-like protease precursor
Sopen01g034720	-8.23	5.90E-15	-10.5	5.70E-21	Subtilisin-like serine protease SBT4

Sopen08g003620	-3.86	6.40E-12	-4.11	7.10E-07	Subtilisin-like protease SBT1.7
Sopen10g033760	-2.98	9.70E-05	-8.04	3.10E-08	Subtilisin-like protease SBT1.7
Sopen01g034710	-4.13	6.80E-03	-9.72	4.50E-31	Subtilisin-like protease SBT1.7
Sopen12g030670	-1.13	7.60E-03	-1.76	2.00E-06	Subtilisin-like protease SBT1.7
Sopen10g033770	-1.77	1.80E-02	-3.70	4.90E-09	Subtilisin-like protease SBT1.7
Sopen08g003550	1.02	1.70E-07	1.18	1.80E-04	Subtilisin-like protease SBT1.7
Sopen07g001240	-5.34	1.00E-09	-5.30	8.90E-07	Chitotriosidase-1 (chitinase activity)
Sopen01g040940	-3.41	2.80E-04	-6.03	1.20E-08	Wound-induced protein WIN1-like (chitin binding)
Sopen09g033970	-5.43	1.70E-14	-3.95	3.60E-06	Pathogenesis-related protein STH-2
Sopen01g048950	-3.72	4.20E-05	-4.92	2.30E-05	Pathogenesis-related protein 1A-like
Sopen10g019020	-4.16	3.80E-07	-3.42	4.80E-12	Kirola-like defense response protein
Sopen04g002880	6.04	1.70E-15	4.14	9.50E-05	Kirola-like defense response protein
Amine oxidase-related defense system					
Sopen06g005510	7.52	2.60E-42	6.46	1.90E-17	Glutamine synthetase catalytic domain
Sopen07g017370	7.54	3.20E-47	6.86	3.90E-16	Copper-containing amine oxidase
Sopen11g004040	7.30	9.20E-15	6.89	3.20E-18	Protein FLOWERING LOCUS D (flavin-containing amine oxidase)
Sopen11g004050	6.04	2.30E-14	4.53	1.20E-15	Protein FLOWERING LOCUS D (flavin-containing amine oxidase)
Sopen01g035050	7.84	5.00E-57	6.07	6.60E-13	Aldehyde oxidase
Sopen05g003410	6.95	2.00E-22	6.11	6.00E-09	Extensin-like protein
Sopen01g001850	2.97	2.10E-10	2.67	1.00E-06	Extensin-like protein

1 Log<sub>2</sub>FC and FDR indicate log<sub>2</sub> (fold-change) and false discovery rate (*P* values adjusted for  
2 multiple-testing), respectively. Positive and negative log<sub>2</sub>FC values indicate higher and lower  
3 expression levels, respectively in high-acylsugar-producing accessions. We placed  
4 *Sopen12g026980* (marked with an asterisk) with other DEGs putatively encoding glutathione S-  
5 transferase; however, this gene is strongly co-expressed with acylsugar metabolic genes (see  
6 Figure 7).

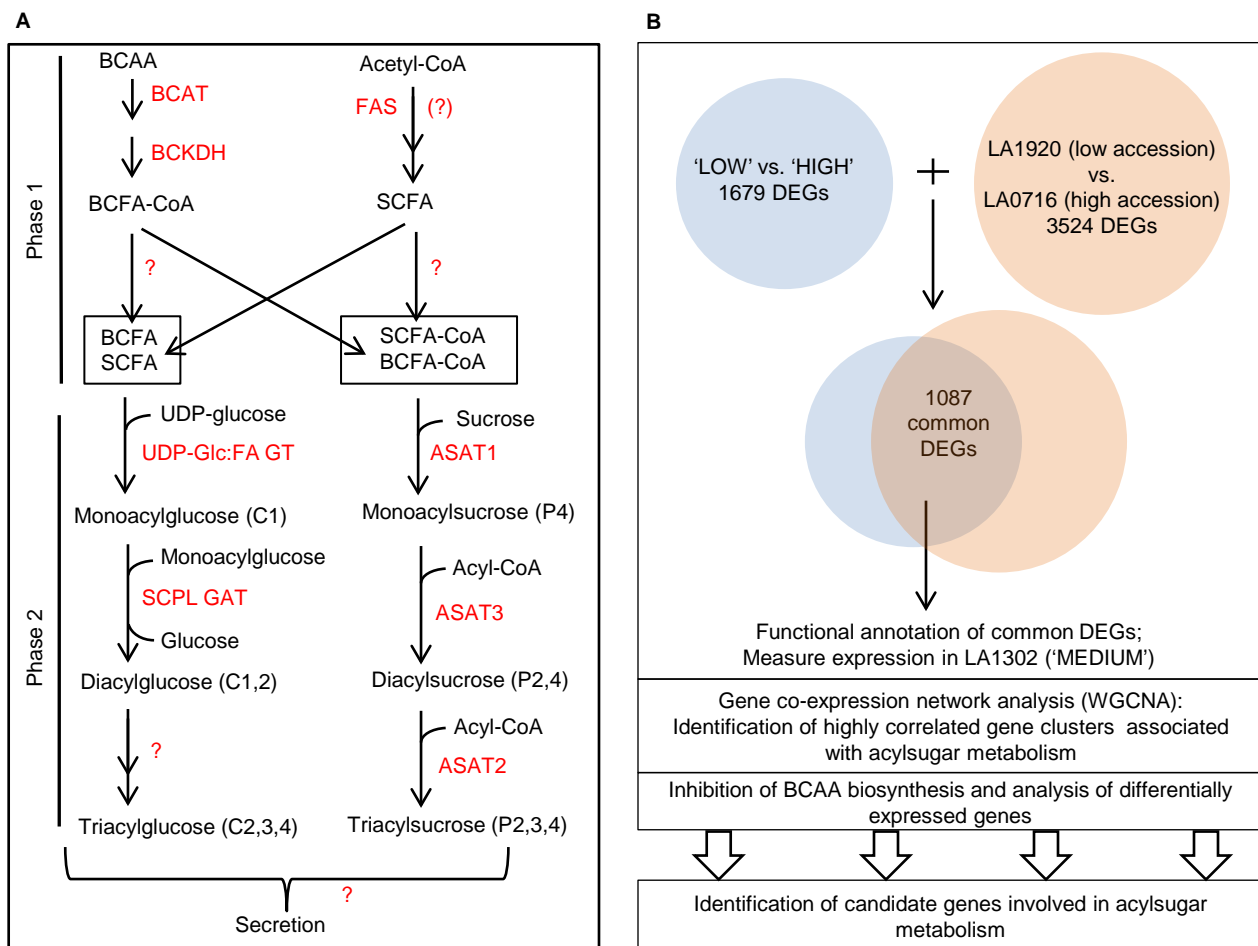
7

8 **Table 5.** Response of selected genes to imazapyr treatment.

Gene ID	Imazapyr 0.1 mM		Imazapyr 1 mM		Annotation
	Log <sub>2</sub> FC	FDR	Log <sub>2</sub> FC	FDR	
Genes related to phase 1 of acylsugar biosynthesis (Table 1 genes)					
Sopen11g004560	0.40	3.35E-01	-2.22	8.69E-12	Acetolactate synthase
Sopen07g027240	0.92	3.20E-03	1.00	2.95E-04	Ketol-acid reductoisomerase
Sopen05g032060	-0.08	8.46E-01	-1.22	1.32E-06	Dihydroxy-acid dehydratase
Sopen08g005060	2.03	1.45E-10	1.91	1.68E-10	Isopropylmalate synthase
Sopen08g005140	-3.11	6.11E-06	-5.63	7.20E-12	Isopropylmalate synthase
Sopen04g030820	-2.90	5.88E-08	-3.73	1.11E-12	Branched-chain aminotransferase-2
Sopen04g026270	-0.70	7.42E-02	3.17	1.18E-21	Branched-chain keto acid dehydrogenase E1 subunit
Sopen01g028100	-0.72	2.61E-02	1.91	1.35E-12	Branched-chain keto acid dehydrogenase E2 subunit
Sopen07g023250	-2.65	2.50E-05	-7.04	1.92E-19	3-Hydroxyisobutyryl-CoA hydrolase
Sopen05g023470	-0.68	3.24E-01	-1.41	8.99E-03	3-Hydroxyisobutyryl-CoA hydrolase
Sopen12g032690	-1.28	4.23E-02	-4.52	9.34E-10	Mitochondrial acyl-CoA thioesterase
Sopen12g004240	-1.55	1.88E-06	-2.50	8.84E-16	KAS IV/ KAS II-like domain
Sopen08g002520	-2.29	9.29E-08	-4.03	7.27E-20	Beta-ketoacyl-ACP synthase III
Sopen05g009610	-2.81	3.03E-05	-4.89	1.08E-11	Beta-ketoacyl-ACP reductase
Sopen12g029240	-2.25	5.40E-05	-3.91	4.51E-12	Enoyl-ACP reductase domain
Sopen02g027670	-2.29	2.18E-04	-4.73	6.94E-13	Acyl-activating enzyme 1
Sopen02g027680	-1.83	2.71E-04	-4.95	1.78E-19	Acyl-activating enzyme 1
Sopen07g023200	-2.28	4.84E-05	-4.86	1.15E-14	Acyl-activating enzyme 1
Sopen07g023220	-2.64	2.13E-06	-6.08	4.25E-20	Acyl-activating enzyme 1
Genes related to phase 2 of acylsugar biosynthesis (Table 2 genes)					
Sopen01g049990	1.81	2.59E-06	1.81	4.01E-07	UDP-glucose:fatty acid glucosyltransferase
Sopen10g020280	0.83	4.02E-03	0.06	8.40E-01	Serine carboxypeptidase-like glucose acyltransferase
Sopen12g002290	-3.07	7.59E-06	-6.55	1.65E-14	Acylsugar acyltransferase 1
Sopen04g006140	-1.62	1.23E-02	-5.99	2.64E-15	Acylsugar acyltransferase 2
Sopen11g026960	-1.40	1.69E-03	-4.30	6.51E-22	Acylsugar acyltransferase 3
Sopen05g030120	-1.16	1.33E-04	-1.91	4.26E-12	Acylsugar acylhydrolase 1

Sopen05g030130	-1.40	3.68E-04	-4.53	6.52E-29	Acylsugar acylhydrolase 2
Sopen09g030520	-1.05	9.93E-04	-1.81	4.01E-10	Acylsugar acylhydrolase 3
Sopen04g001210	-2.34	5.74E-11	-6.12	8.82E-50	Carboxylesterase
Other genes related to acylsugar metabolism (Table 3 genes)					
Sopen04g023150	-0.013	9.64E-01	-0.24	5.11E-01	ABC transporter F family member
Sopen01g047950	0.39	5.51E-01	1.26	6.38E-03	ABC transporter G family member
Sopen04g005380	-1.19	3.17E-03	-2.71	1.05E-12	ABC transporter G family
Sopen12g034820	-2.46	1.74E-05	-5.01	4.43E-14	Pleiotropic drug resistance protein 1-like (ABC-G family)
Sopen03g001870	-1.89	3.65E-05	-2.81	1.34E-10	ABC transporter B family
Sopen07g006810	-3.25	1.92E-05	-5.50	4.99E-10	RUBISCO small subunit
Sopen08g020180	-1.42	3.69E-05	-3.33	3.41E-23	NADP-dependent malic enzyme
Sopen03g040490	-1.32	3.35E-04	-3.67	1.96E-23	Invertase (beta-fructofuranosidase)
Sopen01g042050	-2.17	7.50E-03	-2.43	6.64E-04	Sugar transporter ERD6-like
Sopen05g008450	-1.58	8.73E-03	-4.58	2.43E-12	AP2 domain
Sopen12g021250	-2.87	1.93E-04	-1.56	1.74E-02	AP2 domain
Sopen03g036630	-1.89	5.15E-04	-2.95	1.71E-08	AP2 domain
Sopen10g031080	-1.19	3.82E-02	-2.31	8.15E-06	Homeobox associated leucine zipper
Sopen06g024660	-1.22	6.20E-03	-2.32	1.46E-08	Myb-like DNA-binding domain
Sopen08g028640	-1.59	1.02E-02	-2.31	5.08E-05	AP2 domain
Sopen02g021600	-1.24	4.25E-03	-1.80	4.45E-06	Myb-like DNA-binding domain
Sopen01g037680	-1.46	3.33E-02	-3.52	1.82E-07	TCP family transcription factor
Sopen03g037210	-1.88	8.94E-04	-2.49	1.33E-06	Helix-loop-helix DNA-binding domain

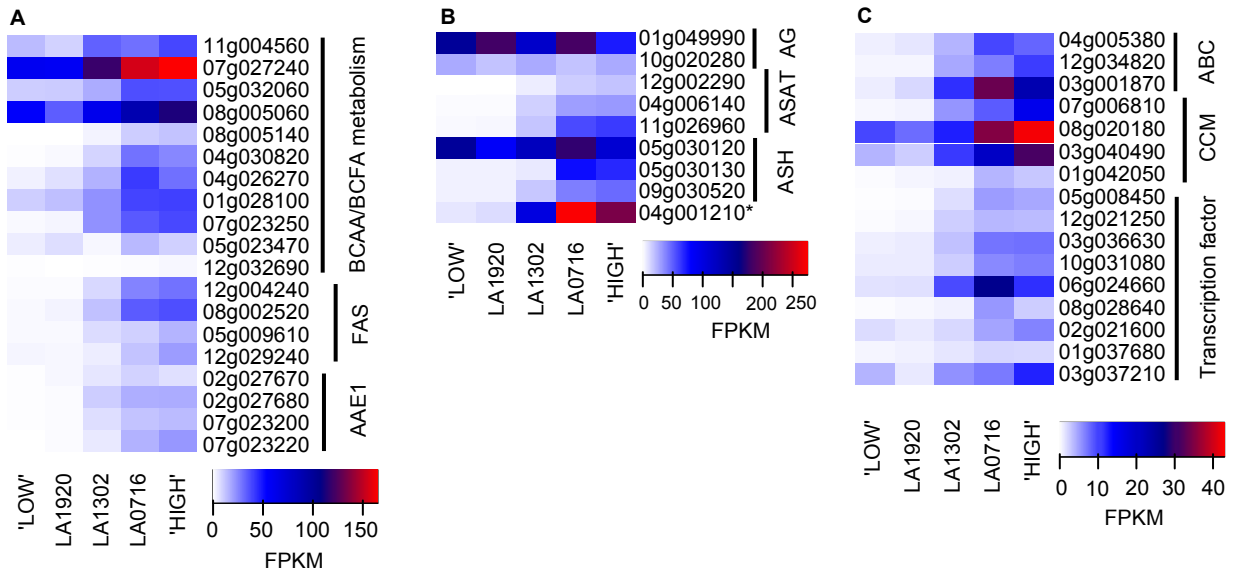
- 1 Log<sub>2</sub>FC and FDR indicate log<sub>2</sub> (fold-change) and false discovery rate (*P* values adjusted for
- 2 multiple-testing), respectively. Positive and negative log<sub>2</sub>FC values indicate higher and lower
- 3 expression levels, respectively in response to the inhibitor treatment compared to control
- 4 solution.



**Figure 1.** Current model of acylsugar production in *Solanum pennellii* and workflow to identify candidate genes for acylsugar metabolism.

**(A)** Branched-chain fatty acids (BCFAs) are derived from branched-chain amino acids (BCAAs), and straight-chain fatty acids (SCFAs) are assumed to be produced by fatty acid synthase (FAS). UDP-glucose:fatty acid glucosyltransferase (UDP-Glc:FA GT) uses free fatty acids to initiate acylglucose biosynthesis, whereas acylsugar acyltransferase 1 (ASAT1) uses fatty acyl-CoAs to initiate acylsucrose biosynthesis. Enzymes are highlighted in red. Unidentified enzymes and transporters are marked with '?'. Double arrows indicate more than one enzymatic step. C1-4 indicates esterification at respective positions on glucose. P2-4 indicates esterification at respective positions on the sucrose pyranose ring.

**(B)** Workflow to identify acylsugar candidate genes. Comparative transcriptomics identified 1679 differentially expressed genes (DEGs) between three low-acylsugar-producing accessions ('LOW') and three high-acylsugar-producing accessions ('HIGH'). An independent comparison with higher sequencing coverage identified 3524 DEGs between another low-acylsugar-producing accession, LA1920, and the high-acylsugar-producing accession LA0716. 1087 DEGs were common to both comparisons. Expression levels of selected candidate genes were determined in LA1302, which accumulates an intermediate level of acylsugars ('MEDIUM'). Weighted gene correlation network analysis (WGCNA) (Langfelder and Horvath, 2008) was performed to identify co-expressed gene networks associated with acylsugar accumulation. Finally, analysis of gene expression in response to inhibition of BCAA biosynthesis narrowed the list of candidates.



**Figure 2.** Heatmaps showing expression levels of genes with known and putative functions in acylsugar metabolism. Genes are designated by their gene identifier numbers (Sopen IDs).

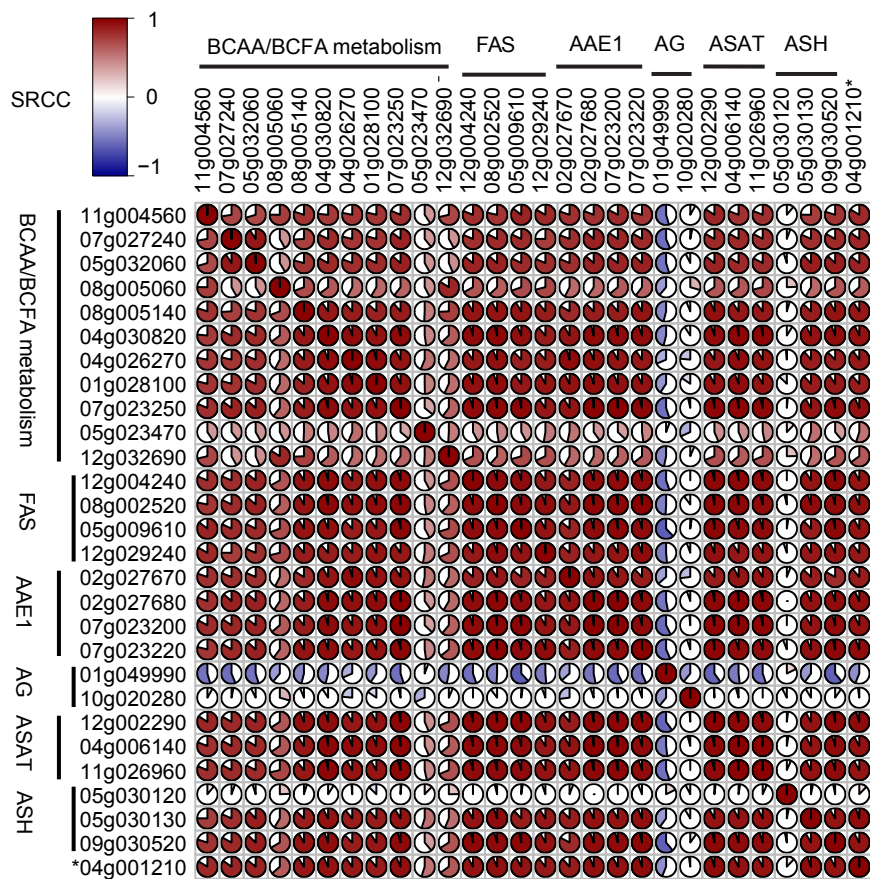
**(A)** Acylsugar phase 1-related genes (Table 1). Expression levels of many candidate genes in LA1302 ('MEDIUM' accession) were intermediate between low- and high-acylsugar-producing accessions.

**(B)** Acylsugar phase 2-related genes (Table 2). *Sopen04g001210* (marked with an asterisk) was identified as a carboxylesterase gene related to ASH, but the encoded protein does not exhibit in vitro activity against tomato acylsucrose fractions (Schillmiller et al., 2016).

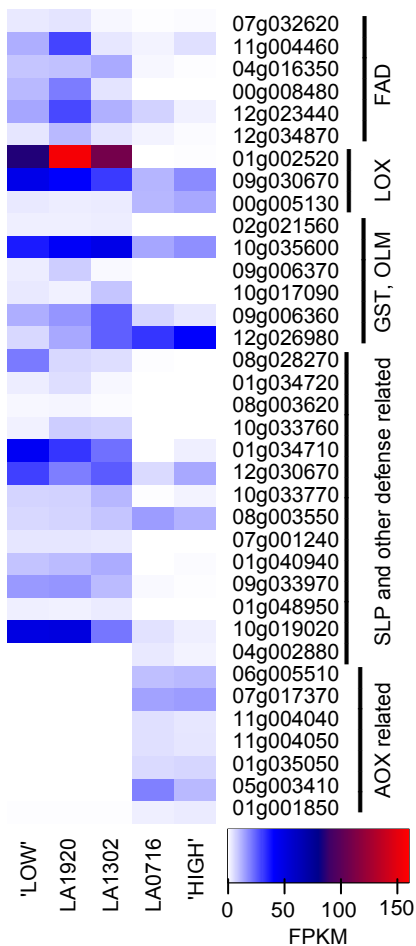
**(C)** Other genes related to acylsugar metabolism (Table 3). Two genes (*Sopen04g023150* and *Sopen01g047950*) were removed from the heatmap due to their very high expression levels. *Sopen04g023150* had FPKM values of 570, 624, 455, 179, and 254 in 'LOW', LA1920, LA1302, LA0716, and 'HIGH' groups, respectively. *Sopen01g047950* had FPKM values of 61, 203, 79, 29, and 19 in 'LOW', LA1920, LA1302, LA0716, and 'HIGH' groups, respectively.

Abbreviations: FPKM= fragments per kilobase of transcript per million mapped reads; BCAA= branched-chain amino acid; FAS= fatty acid synthase components; AAE1= acyl-activating enzyme 1; AG= acylglucose phase 2; ASAT= acylsucrose acyltransferase; ASH= acylsugar hydrolase; ABC= ABC transporters; CCM= central carbon metabolism.

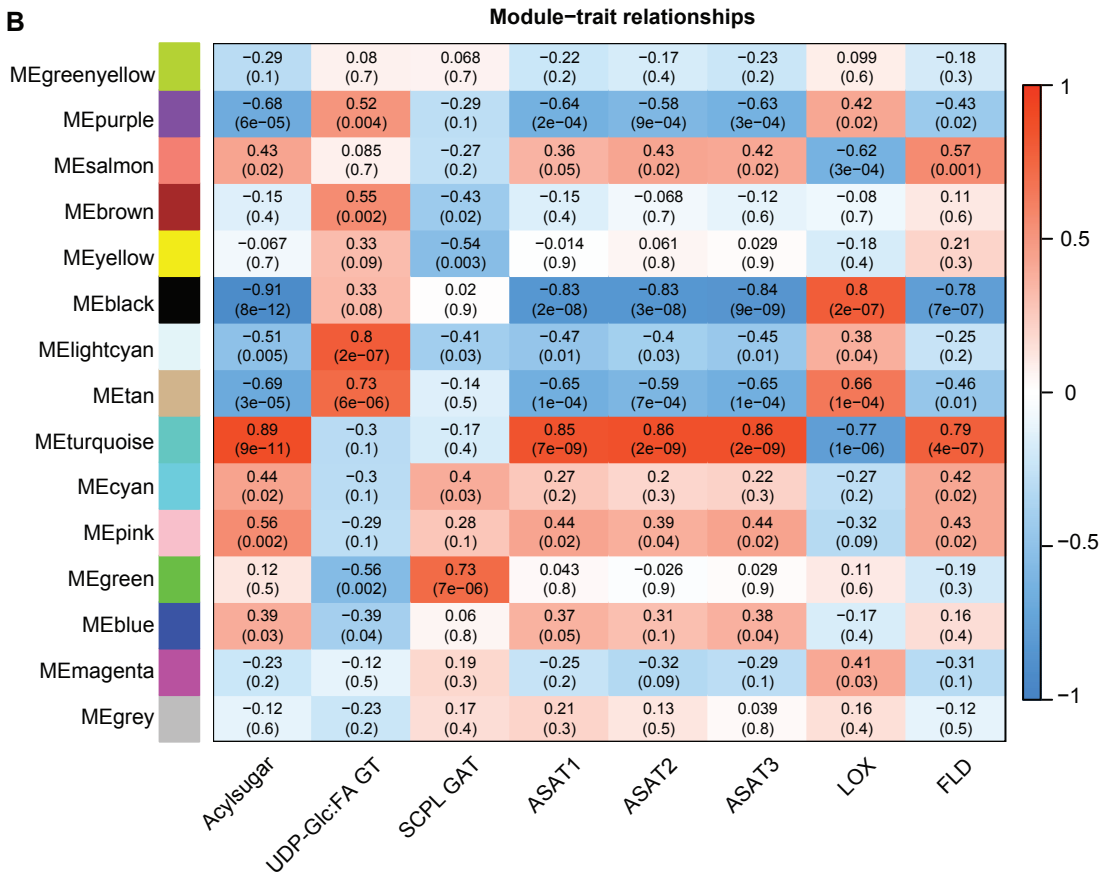
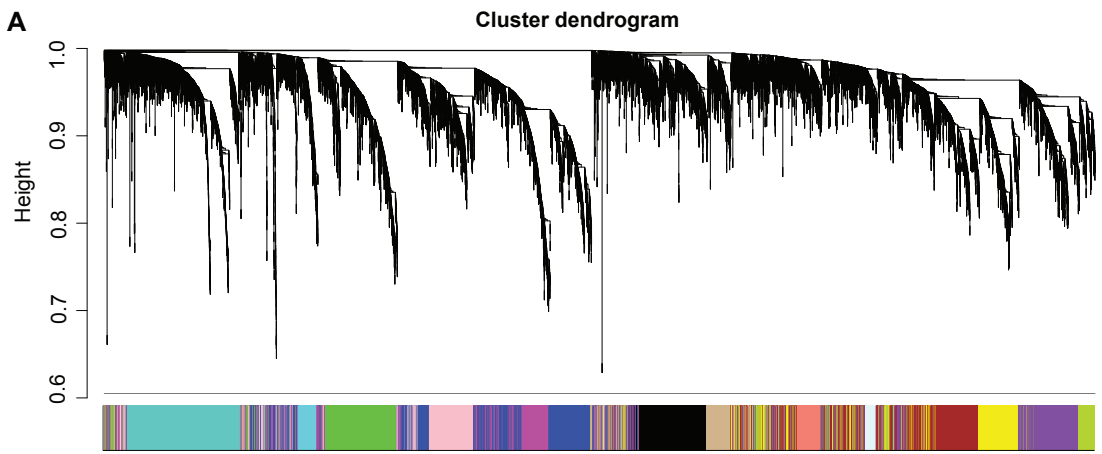




**Figure 3.** Correlation among expression profiles of selected genes. FPKM values for 28 phase 1- and phase 2-related genes (from Table 1 and Table 2, respectively) in our 29 samples were used to determine pairwise Spearman's rank correlation coefficients (SRCC). A [28x28] matrix was created to visualize correlation. SRCC values and associated P values are given in Supplemental Dataset 2. *Sopen04g001210* (marked with an asterisk) was identified as a carboxylesterase gene related to ASH (Schillmiller et al., 2016). Abbreviations: BCAA= branched-chain amino acid; FAS= fatty acid synthase components; AAE1= acyl-activating enzyme 1; AG= acylglucose phase 2; ASAT= acylsucrose acyltransferase; ASH= acylsugar hydrolase.



**Figure 4.** Heatmap showing expression levels of genes putatively involved in plant defense (Table 4 genes). Expression levels of many genes in LA1302 were similar to low-acylsugar-producing accessions but different from high-acylsugar-producing accessions. Some oxylipin-related defense genes, which were upregulated in low-acylsugar-producing accessions, had higher expression levels than amine oxidase-related defense genes, which were upregulated in high-acylsugar-producing accessions. Abbreviations: FPKM= fragments per kilobase of transcript per million mapped reads; FAD= fatty acid desaturase; LOX= lipoxygenase; GST= glutathione S-transferase; OLM= oxylipin metabolism; SLP= subtilisin-like protease; AOX= amine oxidase.

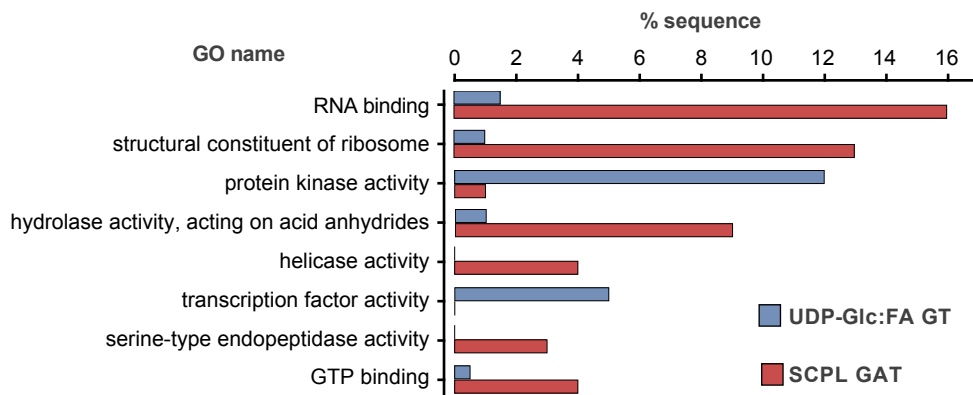


**Figure 5.** Weighted gene correlation network analysis (WGCNA) (Langfelder and Horvath, 2008).

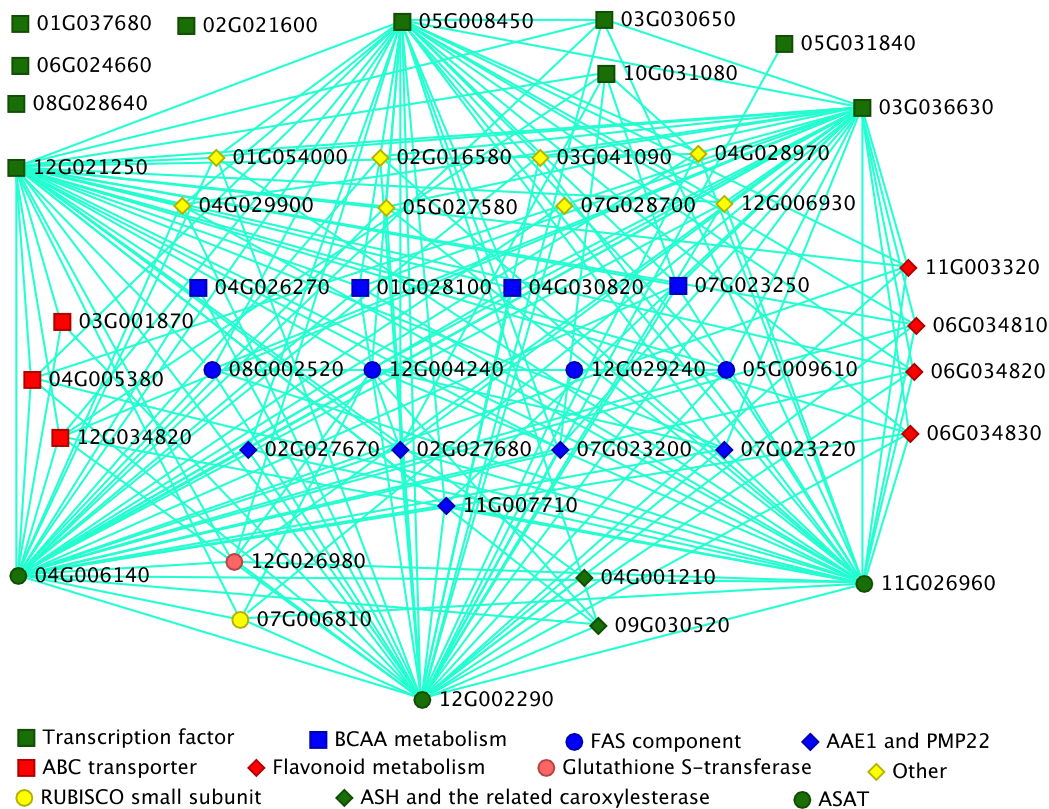
**(A)** Dendrogram clustering 19,378 genes into 15 co-expressed modules based on their expression profiles in all 29 biological samples (see Methods section). For each module, a module eigengene was selected by WGCNA as representative of the expression profile for that module.

**(B)** Association of module eigengenes (MEs) with acylsugar level and seven selected gene expression profiles. The seven genes encode acylglucose biosynthetic enzymes (UDP-Glc:FA GT and SCPL GAT), acylsucrose biosynthetic enzymes (ASAT1, ASAT2, and ASAT3), and defense related proteins (LOX and FLD as representatives of oxylipin-related and amine oxidase-related defense networks, respectively). Spearman's rank correlation coefficient (SRCC) values between MEs and traits are shown; P values associated with SRCC are given in parentheses.

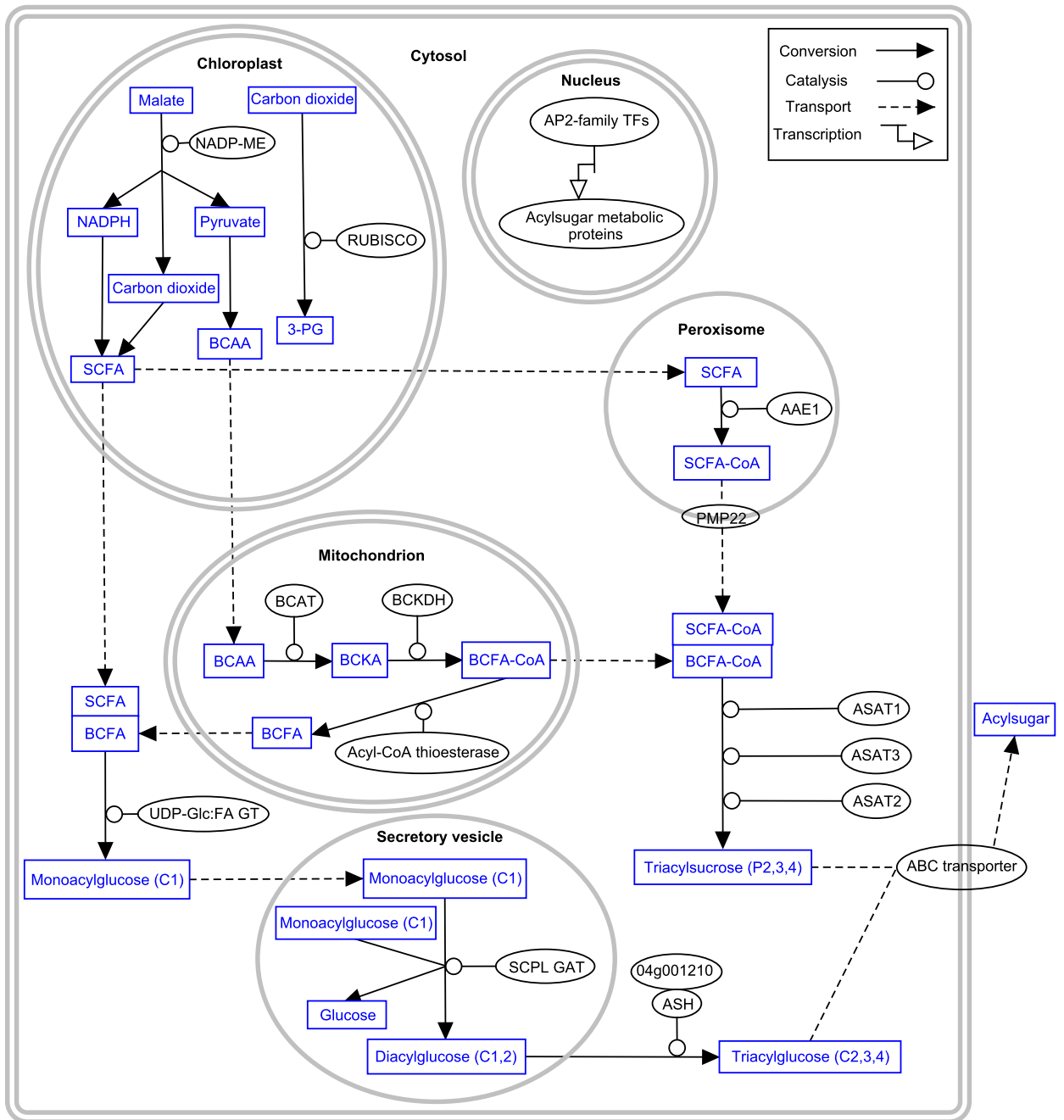
Abbreviations: UDP-Glc:FA GT= UDP-glucose:fatty acid glucosyltransferase; SCPL GAT= serine carboxypeptidase-like glucose acyltransferase; ASAT= acylsugar acyltransferase; LOX= lipoxygenase; FLD= FLOWERING LOCUS D.



**Figure 6.** Enrichment analysis of Gene Ontology (GO) terms associated with genes most strongly correlated with two genes encoding known acylglucose biosynthetic enzymes (UDP-Glc:FA GT and SCPL GAT). Enrichment analysis was performed with Blast2GO (Conesa et al., 2005) using Fischer's exact test ( $P < 0.05$ ). Top-100 genes associated with *Sopen01g049990* (UDP-Glc:FA GT) in both 'lightcyan' and 'tan' modules, and top-100 genes associated with *Sopen10g020280* (SCPL GAT) in 'green' module were used for the analysis. Categories are listed from top to bottom based on their P values (from  $P = 3.57E-06$  for 'RNA binding' to  $P = 4.39E-02$  for 'GTP binding'). GO terms in 'molecular function' subsection only were selected, and redundant categories were combined. A full list of GO terms and associated P values are given in Supplemental Figure 8.



**Figure 7.** Simplified acylsugar metabolic gene network. Nodes and edges represent genes and intramodular connectivities, respectively. We used WGCNA to identify the top-50 genes most strongly connected to each of the eight transcription factor genes (all green squares except *Sopen03g030650* and *Sopen05g031840*) that responded to imazapyr and were in the 'turquoise' module, as well as three acylsucrose biosynthetic genes (three ASATs). These 11 top-50 lists contained only 160 different genes. Using VisANT (Hu et al., 2005), with an edge weight cutoff value of 0.20, 41 genes were identified as most strongly connected. Three genes putatively encoding AP2-family transcription factors (*Sopen05g008450*, *Sopen12g021250*, and *Sopen03g036630*) showed strongest connectivity with acylsugar (and also flavonoid) metabolic genes. This network also includes two additional transcription factor genes (*Sopen03g030650* and *Sopen05g031840*). *Sopen03g030650* was downregulated only at 1 mM imazapyr treatment, whereas *Sopen05g031840* was slightly upregulated at both 0.1 and 1 mM imazapyr treatment.



**Figure 8.** Proposed model of acylsugar metabolism in *S. pennellii*. Role of different subcellular compartments are shown. Metabolites are highlighted in blue within boxes. Enzymes, transporters and transcription factors (TFs) with known and putative functions in acylsugar metabolism are highlighted in black within ovals. Single arrow does not necessarily represent a single enzymatic step. This figure was prepared using WikiPathways (Pico et al., 2008). Abbreviations: NADP:ME= NADP-dependent malate dehydrogenase; 3-PG= 3-phosphoglycerate; BCKA= branched-chain keto acid.