GENETIC ANALYSIS OF ACYLSUGAR METABOLISM IN WILD TOMATO *SOLANUM PENNELLII*

A Dissertation

by

SABYASACHI MANDAL

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Thomas McKnight
Committee Members,	Wayne Versaw
	Alan Pepper
	Patricia Klein
Head of Department,	Thomas McKnight

December 2019

Major Subject: Biology

ABSTRACT

Plant metabolites are sub-divided into two classes- primary and secondary or specialized metabolites. Acylsugars are specialized metabolites secreted by many members of the nightshade family (Solanaceae). These compounds are branched-chain and straight-chain fatty acids esterified to glucose or sucrose. Acylsugars have important roles in plant defense and potential commercial applications. However, several acylsugar metabolic genes remain unidentified, and little is known about regulation of this pathway. Using transcriptomics approaches, acylsugar metabolic genes were identified in a wild tomato species (*Solanum pennellii*). Differential gene expression analyses between low- and high-acylsugar-producing accessions of *Solanum pennellii* revealed that expression levels of known and novel candidate genes (putatively encoding beta-ketoacyl-(acyl-carrier-protein) synthases, peroxisomal acyl-activating enzymes, ABC transporters, and central carbon metabolic proteins) were positively correlated with acylsugar accumulation, except two genes previously reported to be involved in acylglucose biosynthesis. Genes putatively encoding oxylipin metabolic proteins, subtilisin-like proteases, and other antimicrobial defense proteins were upregulated in low-acylsugar-producing accessions, possibly to compensate for diminished defense activities of acylsugars. Transcriptome analysis after biochemical inhibition of biosynthesis of branched-chain amino acids (precursors to branched-chain fatty acids) by imazapyr showed concentration-dependent downregulation of known and most acylsugar candidate genes, but not defense genes. Weighted gene correlation network

analysis identified separate co-expressed gene networks for acylsugar metabolism (including six transcription factor genes and flavonoid metabolic genes) and plant defense (including genes putatively encoding NB-ARC and leucine-rich repeat sequences, protein kinases and defense signaling proteins, and previously mentioned defense proteins). Acylsugars are secreted through trichomes, and virus-induced gene silencing of two trichome preferentially expressed candidate genes for straight-chain fatty acid biosynthesis confirmed their role in acylsugar metabolism. This work will serve as a valuable genetic resource for increasing our understanding of acylsugar metabolism and defense in *Solanum pennellii* and other plants.

DEDICATION

To my beloved parents.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. McKnight, and my committee members, Dr. Versaw, Dr. Pepper, and Dr. Klein, for their guidance and support throughout the course of this research. I am very grateful to Dr. McKnight for not only teaching me how to do science but also for making me a better human being.

I sincerely appreciate all the help from my labmate, Wangming Ji. I am obliged to the Department of Biology faculty, students, and staff for giving me the opportunity to be a part of this great department. I am thankful to Drs. Shovon Mandal, Shahriar Kibriya, Naureen Fatema, Arnab Chatterjee, and Pallavi Mukherjee for their brotherly love. Thanks also go to my friends Dibyendu, Abhinav, Sandeep, Pradeep, and Vrijesh for their support and encouragement. I am indebted to my Bryan-College Station friends Nikita, Nandhini, Sravani, Paola, Yi, Krishna, Ajit, Jitender, Tathagata, Vishal, and Leena Di for making my time at Texas A&M University a great experience.
Finally, special thanks to my parents, grandparents, uncle, and aunt Manju for their encouragement, support, and love.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Dr. Thomas McKnight [advisor], Dr. Wayne Versaw, and Dr. Alan Pepper of the Department of Biology and Dr. Patricia Klein of the Department of Horticultural Sciences.

The primers for qRT-PCR analyses were designed by Wangming Ji, a graduate student in the McKnight lab. Some of the RNA samples were also prepared by him. RNA-Sequencing libraries were constructed and sequenced by members of the Texas A&M Genomics and Bioinformatics center. Acylsugar acyl chain profile analyses were conducted in part by Yohannes Rezenom of the Department of Chemistry.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by a teaching assistantship from Texas A&M University and, in part, from a research grant funded by the US Department of Agriculture (2011-38821-30891).

NOMENCLATURE

RNA-Seq	RNA-Sequencing
NGS	Next-generation sequencing
BCFA	Branched-chain fatty acid
SCFA	Straight-chain fatty acid
BCAA	Branched-chain amino acid
FAS	Fatty acid synthase
UDP-Glc:FA GT	UDP-glucose:fatty acid glucosyltransferase
SCPL GAT	Serine carboxypeptidase-like glucose acyltransferase
UTR	Untranslated region
ASAT	Acylsugar acyltransferase
WGCNA	Weighted gene correlation network analysis
VIGS	Virus-induced gene silencing
TRV	Tobacco rattle virus
FDR	False discovery rate
GO	Gene ontology

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

1. INTRODUCTION, LITERATURE REVIEW, AND PRELIMINARY WORK

1.1. Primary and Secondary Metabolism

There are nearly 400,000 recognized plant species on earth (Royal Botanic garden, Kew), and each species produces a plethora of metabolites. Plant metabolites have been traditionally sub-divided into two classes- primary or central metabolites and secondary or specialized metabolites. Primary metabolites are vital for the survival of organisms and constitute more ancient and evolutionary conserved metabolic pathways, such as photosynthetic carbon fixation, glycolysis, TCA cycle, and amino acid metabolic pathways. On the other hand, specialized metabolites are defined as metabolites that are produced in some species, but not in others; therefore, these compounds are not vital for the survival of most organisms, but do have important functions in growth and development. Due to relaxed selection pressure, specialized metabolism exhibits much greater diversity. For example, there are fewer than 10,000 primary metabolites, whereas the number of specialized metabolites is estimated at roughly 200,000 (Dixon and Strack, 2003). This number is probably an underestimation because specialized metabolites have been characterized from only a handful of plant species. Specialized metabolites are not generally produced in high concentrations, and the efficacy of extraction and detection techniques also contribute to many low-molecular-weight specialized metabolites being unidentified (Pichersky and Lewinsohn, 2011). Additionally, composition of specialized metabolites typically varies within and between species, and biosynthesis of these compounds is often under spatial and/or temporal

regulation, making it hard to estimate the extent of diversity in specialized metabolism (Kroymann, 2011).

How did this tremendous metabolic diversity originate? In order to answer the question, we need to understand the evolution of secondary metabolism. The building blocks of secondary metabolism are derived from primary metabolism. For example, phenylpropanoids (hydroxycinnamic acids, lignins, and flavonoids) are derived from the aromatic amino acid phenylalanine; isopentenyl pyrophosphates give rise to terpenoids; polyketides are synthesized from acetyl-CoA and malonyl-CoA; alkaloids are derived from amino acids, such as tryptophan, tyrosine, arginine, and lysine (Modolo et al., 2009). Many reactions and enzymes of secondary metabolism exhibit close similarities with those of primary metabolism, and gene duplication followed by mutations (neofunctionalization) is an important driver of evolution of specialized metabolites (Modolo et al., 2009). In addition to gene duplication and mutation, novel gene functions for specialized metabolism can originate through a variety of other mechanisms, such as transposition, alteration of gene structure (domain/exon shuffling), and/or through spatial/temporal regulation of gene expression (Pichersky and Lewinsohn, 2011). All these factors contribute to the metabolic diversity of plant specialized metabolism.

1.2. Specialized Metabolism and Recent Research

Many specialized metabolites have important functions in the interaction between plants and their environment, including roles as protectors against harmful UV light, as defense compounds in providing protection against herbivores and microbial

pathogens, as pollinator and/or seed disperser attractants, and as communication compounds (allelopathy and legume-rhizobia interactions) (Kroymann, 2011; Moghe and Last, 2015). These compounds have also provided humans with many beneficial uses, such as pharmaceuticals and cosmetics, biocontrol agents, dyes, and spices (Balandrin et al., 1985). In the past three decades, considerable research has focused on elucidating genetic networks involved in specialized metabolic pathways, such as phenylpropanoids, flavonoids, alkaloids, terpenoids, and glucosinolates. Before the advent of 'omics'-based tools, establishing relationships between specialized metabolic diversity and underlying genes was a big challenge. With tremendous advances in sequencing technologies in the past one or two decades, considerable progress has been made on identifying these genes. Although genome and transcriptome sequences have been made available initially for model plants only, underlying principles from those genomics and transcriptomics studies have been applied later to non-model plants (Kroymann, 2011; Gongora-Castillo et al., 2012).

1.3. Exploiting Natural Genetic Variation as a Tool to Identify Genes Involved in Specialized Metabolism

Natural genetic variation can lead to quantitative and/or qualitative variation, although these two variations are often linked with each other (Owen et al., 2017). Exploiting natural variation in the accumulation of specialized metabolites among different accessions, different studies have identified genes involved in the biosynthesis and regulation of specialized metabolic pathways, such as glucosinolate and terpenoid

pathways (Kliebenstein et al., 2001b; Kliebenstein et al., 2001a; Chen et al., 2003; Tholl et al., 2005). Natural genetic variation can also be exploited to produce experimental genetic materials, such as recombinant inbred lines and near isogenic lines in order to identify metabolic genes (Torjek et al., 2006; Torjek et al., 2008). Genomics and transcriptomics approaches can be used to exploit natural genetic variation and subsequently identify genes associated with a particular trait. For example, genome sequence information from different species or different accessions from one species can be used in comparative genomics studies to identify single nucleotide polymorphisms useful for quantitative trait locus (QTL) mapping and genome-wide association studies (Kroymann, 2011). Genome-wide polymorphism information has been used in several studies to identify specialized metabolic genes, including some of the genes previously identified by conventional QTL mapping (Nordborg et al., 2005; Atwell et al., 2010; Chan et al., 2010; Li et al., 2010; Cao et al., 2011; Li et al., 2014).

Transcriptomics approaches have also been used extensively to identify specialized metabolic genes, since differential gene expression of underlying genes is mostly responsible for quantitative variation in specialized metabolism (Wentzell et al., 2007). To identify genes for missing steps in a metabolic pathway, gene co-expression network analysis can also be performed, if one or more characterized genes are known to be involved in a pathway. This approach has been used to identify genes involved in glucosinolate metabolism in the Brassicaceae (He et al., 2009; Knill et al., 2009; Sawada et al., 2009; Albinsky et al., 2010). Similarly, transcription factors involved in the regulation of glucosinolate metabolism have also been identified using transcriptomics

approaches (Gigolashvili et al., 2007; Hirai et al., 2007; Sonderby et al., 2007; Gigolashvili et al., 2008; Sonderby et al., 2010). Production of specialized metabolites such as plant defense compounds is often triggered in response to elicitor treatment or pathogen attack. Transcriptome analysis in response to methyl jasmonate (an elicitor) treatment led to the identification of genes involved in alkaloid biosynthesis in *Catharanthus roseus* (Giddings et al., 2011; Geu-Flores et al., 2012). Similarly, biosynthesis of etoposide in *Podophyllum hexandrum* can be induced by wounding, and the ten-step biosynthetic pathway was elucidated through co-expression network analysis (Lau and Sattely, 2015). The use of transcriptome data and co-expression network analysis is particularly useful for non-model plants where complete genome sequence is still unavailable, or for plants with intractable genomes (such as ~16 gigabase genome in *Podophyllum hexandrum*). The main objective of this work was to identify candidate genes involved in acylsugar biosynthesis using transcriptomics approaches. For this purpose, differential gene expression analyses were performed to exploit natural genetic variation in acylsugar production among different accessions of a wild tomato species (*Solanum pennellii*). A chemical genetics approach was also taken using a biochemical inhibitor of acylsugar precursor biosynthesis. Additionally, gene co-expression network analysis was performed using gene expression data from various samples to identify candidate gene networks involved in acylsugar metabolism. Virus-induced gene silencing of two genes in this network confirmed the validity of our approach.

1.4. RNA-Seq as a Tool of Transcriptomics Studies

RNA-Sequencing (RNA-Seq) is a tool that can determine the sequence and relative abundance of RNA molecules in biological samples using next-generation sequencing (NGS) technologies. Before the advent of NGS technologies, gene predictions and limited EST databases were used to study the transcriptome of a biological sample, thereby giving partial and biased results. With extremely high levels of sensitivity and accuracy, RNA-Seq analyses are capable of giving the complex landscape and dynamics of any biological sample's transcriptome, including large and small RNAs, abundant and rare transcripts, novel transcripts, splicing isoforms, and gene-fusion transcripts (Martin and Wang, 2011). However, commonly used NGS technologies, such as Illumina, SOLiD, 454, etc. give short (50-500 bp) nucleic acid sequences, which require effective and accurate assembly programs to reveal the complex transcriptome (except for small RNA molecules such as miRNA, siRNA, snRNA, piRNA, etc.), if a complete genome sequence is not available for a species (or closely related species). Transcriptome assembly from short reads is similar to genome assembly except for three major differences (Martin and Wang, 2011). First, the sequencing depth of DNA generally remains the same across a genome, whereas the sequencing depth of transcripts differs greatly. Second, transcriptome assembly may be strand-specific whereas genome assembly generally uses both strands. Third, different RNA molecules can be generated from the same locus (alternative splicing) thereby creating ambiguity in the transcript assembly process. However, recent advances in

transcriptome assembly programs, such as Trinity, Velvet/Oases, SOAPdenovo-Trans, and Trans-ABySS, have made it possible to address these challenges (Vijay et al., 2013).

Typical mRNA-Seq transcriptome analysis experiments generally follow these steps: (i) isolation of total RNA and enrichment of mRNA using oligo-dT columns or hybridization-based rRNA depletion methods, (ii) preparation of cDNA (reverse transcription of mRNAs using either random primers or oligo-dT primers) and sequencing libraries, (iii) NGS of cDNA libraries, (iv) filtering low-quality reads and removing sequencing artifacts, (v) assembly of transcripts using short-read assembly programs, (vi) functional annotation of the assembled transcripts and (vii) 'expression counting (mapped read counting)' to identify differentially expressed genes in two or more samples. These steps are usually applied for RNA-Seq analyses in non-model plants where complete genome sequences are unavailable. For plants with available complete genome sequences, steps (v) and (vi) can be skipped; instead, filtered high-quality reads are mapped to the genome, followed by 'expression counting (mapped read counting)' to perform differential gene expression analyses or co-expression analyses.

1.5. Importance of Acylsugar Metabolism in the Solanaceae

The nightshade family (Solanaceae) includes some of world's most important crops, such as potato (*Solanum tuberosum*), tomato (*Solanum lycopersicum*), eggplant (*Solanum melongena*), and peppers (*Capsicum* spp.). Medicinal plants such as *Datura* spp., *Mandragora* spp., and *Atropa belladonna* also belong to this family. Tobacco (*Nicotiana* spp.) and the ornamental plant petunia are also used as model plants to

investigate fundamental biological processes. One of the most extensively studied specialized metabolic pathways in the Solanaceae is the acylsugar pathway.

Many solanaceaous plants secret acylsugars through glandular trichomes. These aerially secreted metabolites have important roles in plant defense. For example, acylsugars from the wild tomato *Solanum pennellii* exhibit antifeedant and antiovipositional properties against insect pests and also exert toxic effects on different herbivores (Hawthorne et al., 1992; Juvik et al., 1994; Liedl et al., 1995). Acylsugars from other members of the Soalanaceae also have important roles in providing protection against herbivores and plant pathogens (Chortyk et al., 1997; Hare, 2005; Luu et al., 2017). For example, acylsugars from *Nicotiana attenuata* are often metabolized by hornworm (*Manduca sexta*) and these worms release volatile organic compounds that attract predatory ants (Weinhold and Baldwin, 2011). Additionally, amphipathic acylsugar molecules provide an additional source of water for the plants by reducing the surface tension of adsorbed dew (Fobes et al., 1985). Acylsugars also have potential applications as pesticides (Puterka et al., 2003), food additives (Hill and Rhode, 1999), cosmetic and personal care products (Hill and Rhode, 1999), antibiotics (Chortyk et al., 1993), and anti-inflammatory compounds (Perez-Castorena et al., 2010). The efficiency of acylsugar insecticidal and miticidal activities can sometimes exceed that of a comparable insecticidal soap (Puterka et al., 2003). Due to these important biological roles and potential commercial applications of acylsugars, a detailed understanding of the genetic network involved in acylsugar metabolism is required for successful crop breeding programs and metabolic engineering of acylsugar production. In this work,

wild tomato *S. pennellii* was used as a model to identify genes involved in acylsugar metabolism.

1.6. Acylsugars in Wild Tomato *S. pennellii*

S. pennellii (Solanaceae) is a perennial shrub, generally 0.5-1 m tall. This plant species is native to Peru and is an important member of the Loma formations on the coastal deserts of the western Andes. This plant is known for its extremely low water usage for survival, and it grows on dry, rocky and sandy soil below 3000 m altitude (Shapiro et al., 1994). Presumably in order to reduce water loss in these dry habitats and to provide protection against herbivores and pathogens, *S. pennellii* secrets acylsugars through trichomes. Trichomes are generally unbranched, 2-3 celled and are found extensively on the leaf surface, and also on the surface of stems and even fruits. Dead insects and dust are often found on the sticky plant surface indicating that these compounds may also help this plant against herbivory. Other studies indicate that these metabolites have antibacterial and antifungal activities (Rodriguez et al., 1993).

S. pennellii acylsugars are low-molecular-weight glycolipids, with three short-chain fatty acids (C4-C12) esterified to glucose or sucrose (Shapiro et al., 1994). These glycolipids are novel in the plant kingdom in the sense that they are built on a glucose (or sucrose) backbone and secreted, unlike other plant glycolipids that are built on a glycerol backbone and are membrane lipids. *S. pennellii* acylsugars accumulate up to 20% of the leaf dry weight, which is a tremendous metabolic investment (Fobes et al., 1985). These secreted glycolipids can be extracted easily by dipping the plant material

into ethanol. Transesterification of purified acylsugars yields one molecule of glucose and three molecules of short-chain fatty acid esters, which can also be used as biogasoline. This process is analogous to biodiesel production, which yields one molecule of glycerol and three molecules of long-chain (C16 and longer) fatty acid esters.

Biofuels are generally liquid fuels obtained from biomass of living organisms. Bioethanol and biodiesel are two of the most commonly used biofuel types at present. Bioethanol, obtained from a variety of plant species by alcoholic fermentation, has poor energy density. On the other hand, biodiesels, which are esters of long chain fatty acids (generally longer than C16), are produced by the transesterification of oils. These are generally obtained from algae and some plant species e.g. jatropha, rapeseed, mustard, sunflower. However, biodiesel production has some challenges including oil extraction methods, downstream processing, and land and water use, particularly for algae. Biogasoline, with C8 carbon chains, falls into a biofuel gap between C2 ethanol and C18 biodiesel fuels. *S. pennellii*, a wild relative of cultivated tomato, can be used as a novel biofuel feedstock. This non-crop plant can be grown on marginal lands and its secreted glycolipids contain C4 to C12 fatty acids that can be converted into biogasoline easily. *S. pennellii* transesterified products are oxygenated, have low sulfur content, and are predicted to be compatible with current fuel engines. The byproduct of the transesterification reaction (glucose and some sucrose) and the remaining lignocellulosic biomass can be used to produce bioethanol through alcoholic fermentation.

1.7. Acylsugar Biosynthetic Pathway in *S. pennellii*

Acylsugars are aliphatic acyl esters of sugar molecules, such as glucose and sucrose. Tremendous diversity in acyl chain composition exists among different members of the Solanaceae; carbon chain length varies from C2 to C12, and different isomer structures are present (anteiso-, iso-, and straight-chain) (Kroumova et al., 2016). In *S. pennellii*, acylsugars are represented by mostly 2,3,4 tri-*O*-acylglucose molecules with trace amounts of acylsucrose molecules (Shapiro et al., 1994; Schilmiller et al., 2012; Schilmiller et al., 2015). Both branched-chain fatty acids (BCFAs) and straight-chain fatty acids (SCFAs) are esterified to glucose and sucrose molecules. BCFAs are derived from branched-chain amino acids (BCAAs), such as valine, isoleucine, and leucine, whereas SCFAs are presumably derived via a *de novo* fatty acid biosynthetic pathway (Walters and Steffens, 1990). Most abundant BCFAs are 2-methylpropanoic acid, 3-methylbutanoic acid, 2-methylbutanoic acid, and 8-methylnonanoic acid. 2-methylpropanoic acid, 3-methylbutanoic acid, and 2-methylbutanoic acids are derived from valine, leucine, and isoleucine, respectively, via transamination reactions followed by one-carbon removal by a branched-chain dehydrogenase complex. The other abundant BCFA (8-methylnonanoic acid) is generated from 2-methylpropanoic acid via a two-carbon chain-elongation process catalyzed by fatty acid synthase (FAS) complex (Kroumova et al., 1994; van der Hoeven and Steffens, 2000; Kroumova and Wagner, 2003). Most abundant SCFAs are *n*-decanoic acid and *n*-dodecanoic acid, although trace amounts of *n*-octanoic acid is also found in certain accessions (Shapiro et al., 1994).

1.7.1. Acylglucose Biosynthesis in *S. pennellii*

Previous biochemical studies reported that in the first step of *S. pennellii* triacylglucose biosynthesis, a short chain fatty acid reacts with UDP-glucose, and the first acyl group is added to the hydroxyl group of glucose C1 (Ghangas and Steffens, 1993; Kuai et al., 1997). This step is catalyzed by the enzyme UDP-glucose:fatty acid glucosyltransferase (UDP-Glc:FA GT). Two types of chain length-specific glucosyltransferases can participate in this fatty acid activation step: (i) one that activates shorter (C4-C6) chain fatty acids, (ii) the other that activates longer (C8-C12) chain fatty acids (Kuai et al., 1997). The second step of acylglucose biosynthesis involves a disproportionation reaction catalyzed by a serine carboxypeptidase-like glucose acyltransferase (SCPL GAT) (Li et al., 1999; Li and Steffens, 2000). In this step, two molecules of 1-*O*-acylglucose react with each other to form one molecule of glucose and one molecule of 1,2-diacylglucose. Biochemical studies show that this acyltransferase belongs to the serine carboxypeptidase protein family with a highly conserved active site comprised of serine, histidine and aspartate (Li and Steffens, 2000). cDNAs encoding the first two proposed enzymes of the pathway have been cloned. However, according to this model of acylglucose biosynthesis, there are probably two or three more uncharacterized enzymes in the pathway. A third enzyme is hypothesized to add the third acyl group to glucose, since no tri-substituted acylglucose molecules are produced by the second enzyme (SCPL GAT). At some point, the fatty acid attached to C1 must be removed. It is not clear whether this C1 fatty acid is moved to C3 or C4, or whether four fatty acids are attached first and then the one at C1 is subtracted (Figure 1A).

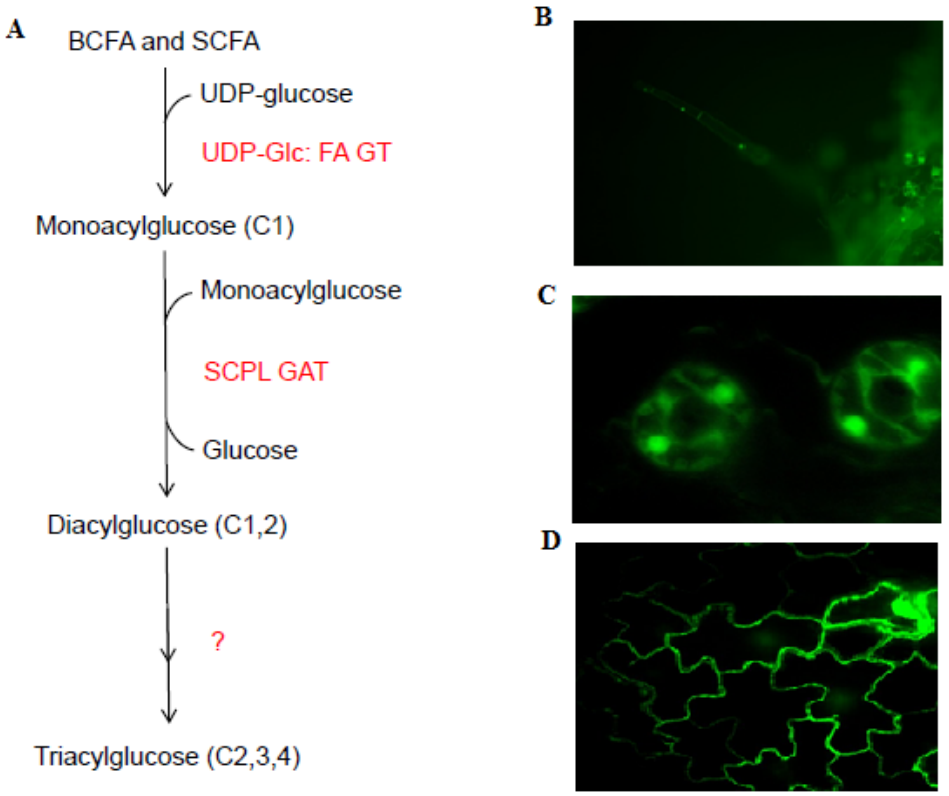

Figure 1. Previous model of acylglucose biosynthesis and localization of the first enzymatic step. (A) Two enzymes (UDP-Glc:FA GT and SCPL GAT; highlighted in red) were reported to initiate acylglucose biosynthesis in *S. pennellii*. Double arrows indicate more than one enzymatic step. (B), (C), and (D) Expression of green fluorescent protein under the control of *UDP-Glc:FA GT* promoter in trichome (B), guard cells (C), and epidermal cells (D).

We cloned and sequenced the two genes encoding UDP-Glc:FA GT and SCPL GAT, respectively. The *UDP-Glc:FA GT* gene does not contain any introns and the open reading frame is 1413 bp long, whereas the *SCPL GAT* gene contains 13 introns and the open reading frame is 1395 bp long. Inverse PCR was used to identify the promoter regions of these two genes. Briefly, genomic DNA was digested with a restriction enzyme (*Eco*RI), the digested fragments were self-ligated to form circular DNA

molecules, and the unknown promoter region was PCR-amplified by using primers within the known open reading frame region. Using this method, we obtained about 3.6 kb and 2 kb of DNA sequences upstream of the translation start site for these two genes, respectively. Next, in order to determine the 5' UTR regions of these two genes, the 5' RACE technique was used (ExactStart^{TM} eukaryotic 5' RACE kit, Epicenter). The 5' UTR regions are about 175 bp and 100 bp long, respectively.

To determine whether expressions of the two genes encoding UDP-Glc:FA GT and SCPL GAT, respectively, are restricted to trichomes, the 3.4 kb and 2 kb promoter regions for these two genes, respectively, were fused to the green fluorescent protein and these reporter constructs were expressed in tobacco. Green fluorescence was detected using a fluorescence microscope in trichomes, and also in epidermal cells, guard cells and mesophyll cells (Figure 1). The observed expression patterns suggest that these two genes are not trichome-specific.

Semi-quantitative PCR showed that transcript levels of the *UDP-Glc:FA GT* and the *SCPL GAT* do not positively correlate with acylsugar amounts in different accessions (McKnight, personal communication). In fact, the *UDP-Glc:FA GT* has higher expression levels in low-acylsugar-producing accessions and in LA4135, an F_1 hybrid (*S. lycopersicum* x *S. pennellii*) that produces negligible amount of acylsugar. Recently, CRISPR-Cas9-mediated gene editing showed that a trichome-expressed invertase is responsible for acylglucose biosynthesis in *S. pennellii* (Leong et al., 2019). Together these results suggest that the UDP-Glc:FA GT and the SCPL GAT are not involved in *S. pennellii* acylglucose biosynthesis.

1.7.2. Acylsucrose Biosynthesis in *S. pennellii*

In cultivated tomato, acylsugars are predominantly acylsucrose molecules, whereas in the wild tomato, acylsugars are mostly acylglucose molecules with some acylsucrose molecules (Shapiro et al., 1994; Schilmiller et al., 2012; Schilmiller et al., 2015). In *S. lycopersicum*, sequential activities of four acylsugar acyltransferases (ASATs) produce tetraacylsucrose (three acyl groups on the pyranose ring and one acyl group on the furanose ring) (Schilmiller et al., 2012; Schilmiller et al., 2015; Fan et al., 2016). The *S. pennellii* ortholog of Sl-*ASAT4* does not encode a functional protein, and therefore only triacylsucrose molecules are found in *S. pennellii* (three acyl groups on the pyranose ring only) (Schilmiller et al., 2012). Acylsucrose biosynthesis is initiated by ASAT1 in both *S. pennellii* and *S. lycopersicum*; however, flipped sequential activities of ASAT2 and ASAT3 in *S. pennellii* are responsible for triacylsucrose biosynthesis (Figure 2) (Fan et al., 2017). ASATs belong to the BAHD family of acyltransferase, and genes encoding these enzymes are expressed in the tip cell of type I/IV trichomes in cultivated tomato (Schilmiller et al., 2012; Schilmiller et al., 2015; Fan et al., 2016).

1.8. The Genome of *S. pennellii*

The genome of *S. pennellii* (accession LA0716) was sequenced using Illumina sequencing technology with ~190-fold coverage (Bolger et al., 2014a). There are 12 chromosomes (n=12), and the assembled genome size is 942 megabase pairs (Mbp), approximately seven times that of *Arabidopsis thaliana*, and slightly larger than the cultivated tomato genome (predicted genome size is approximately 900 Mbp, of which

760 Mbp were assembled) (The Tomoto Genome Consortium, 2012; Fernandez-Pozo et al., 2015). Assembled sequences were compared against publicly available BAC sequences, and an accuracy of >99.9% was reported (Bolger et al., 2014a). Most of the gaps were restricted to pericentromeric regions, and realigning reads provided a satisfactory accuracy of gap-filled regions. 44,966 protein-coding genes (48,923 transcript models) were annotated, with 32,273 high-confidence genes.

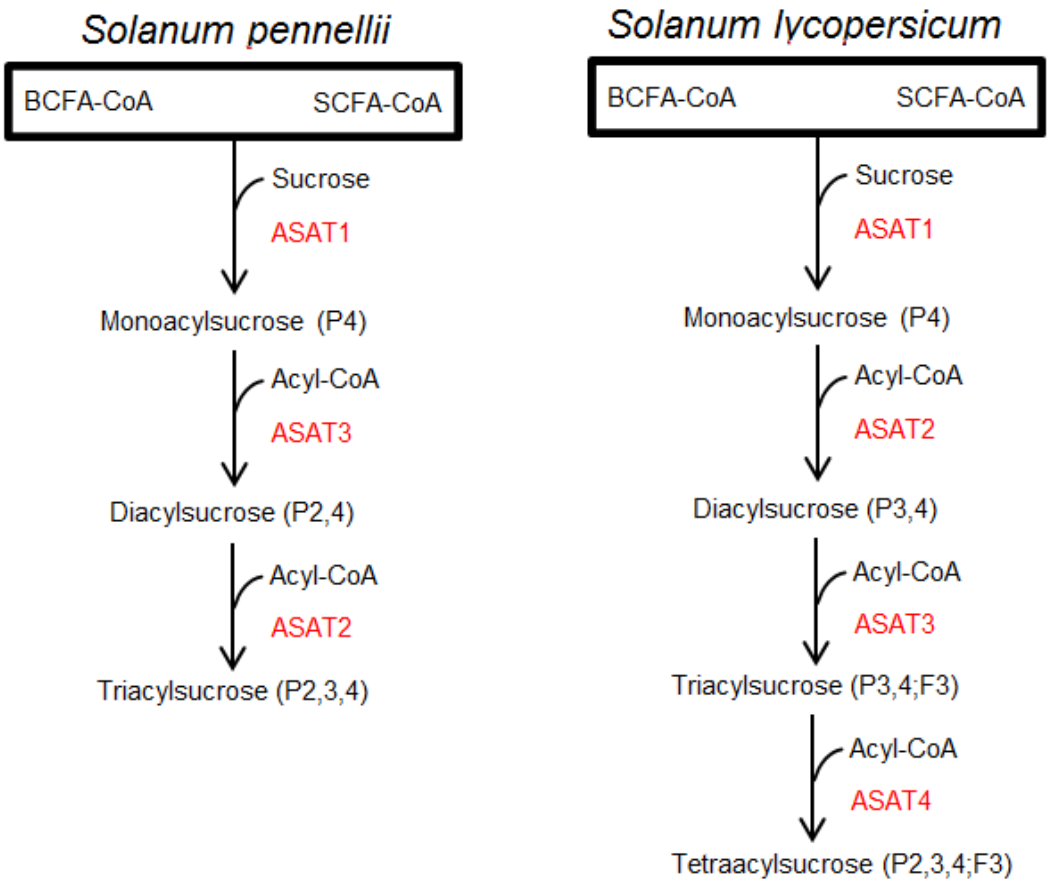

Figure 2. Acylsucrose biosynthesis in *S. pennellii* and *S. lycopersicum*. P2-4 indicates esterification at respective positions on the sucrose pyranose ring. F3 indicates esterification at third position on the sucrose furanose ring. Acylsugar acyltransferase (ASAT) enzymes are highlighted in red.

1.9. Specific Aims

The main objective of this work was to identify and validate candidate genes involved in acylsugar metabolism in the wild tomato *S. pennellii*. To achieve this goal, five specific aims were employed.

Specific Aim1: Construction and characterization of the *S. pennellii* transcriptome.

Specific Aim2: Differential gene expression analysis to identify genes that are differentially expressed between low- and high-acylsugar-producing accessions.

Specific Aim3: Differential gene expression analysis to filter Aim#2 candidate genes that respond to imazapyr (biochemical inhibitor of acylsugar biosynthesis)

Specific Aim4: Gene co-expression network analysis to identify acylsugar metabolic gene networks using gene expression profiles from Aim#2 and Aim#3 experiments.

Specific Aim5: Functional validation of two candidate genes using virus-induced gene silencing methods.

1.9.1. Specific Aim#1: Construction and Characterization of the *S. pennellii* Transcriptome

1.9.1.1. Background Information and Approach

With the advancements in sequencing technology and data analysis methods, combined with decreased cost, NGS technologies are widely used for the construction

and annotation of transcriptomes, which in turn can be used for genetic studies including gene discovery in non-model organisms (Martin and Wang, 2011; Cahais et al., 2012). The first aim of this study was to construct and annotate the *S. pennellii* transcriptome, which can be used as a reference transcriptome for subsequent RNA-Seq experiments. For this purpose, plants were grown under greenhouse conditions and total RNA was extracted from leaf tissues (~10-week old plants) using the Ambion RNAqueous Kit. After DNase treatment of RNA samples (Ambion TURBO DNA-free™ Kit), integrity and quality of RNA samples were checked on an Agilent Bioanalyzer 2100. Poly A^+ RNA molecules were purified using an oligo-dT column and then reverse transcribed to cDNA using random primers. Illumina RNA-Seq libraries were constructed from the cDNA and sequenced using the Illumina HiSeq 2000 platform (paired-end reads, 2x101 bp) at the Texas A&M AgriLife Genomics and Bioinformatics Services center. Approximately 203 million reads were obtained from NGS and these reads were processed for further analysis using the Galaxy platform (open-source, web-based platform) (Giardine et al., 2005; Afgan et al., 2016).

Low quality reads were removed using the Trimmomatic program (version 1.0.0) (Bolger et al., 2014b), and the remaining 173 million reads (average quality ≥ 20, minimum length= 80) were used for transcriptome assembly after quality assessment with the FastQC program (version 0.52) (Andrews, 2010). Initially, transcriptome assembly was performed using the Velvet/Oases *de novo* transcriptome assembler (k-mer values of 21, 27, 33, 39, 45, 51, 57, 63, 69, 75, 81 and merged at k-mer of 51) (Zerbino and Birney, 2008). Transcripts were also assembled with the Trinity assembler

(version 0.3) (Grabherr et al., 2011; Haas et al., 2013), and the results from the two analyses were compared (Table 1). Because Trinity provided a reasonable number of isotigs, we preferred this assembly.

Table 1. Comparison between Oases- and Trinity-assembled transcriptome statistics.

	Oases	**Trinity**
Minimum isotig length (nt)	100	201
Maximum isotig length (nt)	12,081	10,795
Mean isotig length (nt)	748.71	740.07
Standard deviation (nt)	733.93	679.79
Median isotig length (nt)	491	464
N50 isotig length (nt)	1,276	1,151
Number of isotigs	396,878	68,864
Number of isotigs ≥ 1kb	106,371	16,688
Number of isotigs in N50	74,662	13,722
Number of bases in all isotigs	297,146,998	50,964,175
Number of bases in isotigs ≥ 1kb	184,418,914	28,665,284
GC Content of isotigs	40.05 %	39.58 %

Isotigs represent assembled transcript sequences. N50 indicates the length for which the collection of all isotigs of that length or longer contains at least half of the sum of the lengths of all isotigs.

Functional annotation of the transcriptome was performed using the software program Blast2GO (Conesa et al., 2005; Gotz et al., 2008). Gene ontologies were assigned to each transcript. Gene Ontology (GO) is a standardized bioinformatics initiative that offers a dynamically updated, controlled and unified representation system to describe the features of genes and their products (Ashburner et al., 2000). For this, the assembled transcripts were searched against protein databases such as NCBI-NR, Swiss-Prot, KEGG and COG (BlastX with e-value cutoff = 1E-06) and the top 20 hits were

reported. Hits with the best alignment were used for determining the direction of the transcripts. In the case of hits with conflicting results between different protein databases, a priority order was followed (NCBI-NR>Swiss-Prot>KEGG>COG). Transcripts without any alignment were further analyzed using the ESTScan program (Iseli et al., 1999) to determine their coding regions and sequence direction.

Full length open reading frame sequences of the two genes previously reported to be involved in acylglucose biosynthetis (*UDP-Glc:FA GT* and *SCPL GAT*) were identified in the Trinity-assembled transcriptome. To further evaluate the completeness of this transcriptome assembly, 458 Core Eukaryotic Genes Mapping Approach (CEGMA) proteins from *Arabidopsis thaliana* were searched against the database of these assembled transcripts (NCBI tBlastN with e-value cutoff = 1E-08). CEGMA is a tool used to verify enrichment of the transcript population (Parra et al., 2007). In CEGMA analysis, open reading frames of proteins consistently represented in all eukaryotes are searched against the database of assembled transcripts (fasta file). 456 out of 458 proteins gave BLAST hits (132 proteins gave 100% coverage with ≥ 70% identity) indicating the high quality of this assembly.

1.9.1.2. Impact

The genome of *S. pennellii* was published with high-confidence (Bolger et al., 2014a), and it was used as a reference for subsequent RNA-Seq analyses (aim2 and aim3). There are 44,965 gene models, which correspond to 48,923 transcript models, and these numbers support our preference for the Trinity assembly. Our assembled

transcriptome will serve as a useful genetic resource for future studies in the wild tomato.

1.9.2. Specific Aim#2: Gene Expression Analysis to Identify Genes That Are Differentially Expressed between Low- and High-Acylsugar-Producing Accessions

1.9.2.1. Background Information and Approach

About 50 different accessions of *S. pennellii* can be obtained from the Tomato Genetics Resource Center. These accessions differ from each other in terms of the amount of acylsugar secretion (~20% of leaf dry weight to negligible), and they also vary in their acylsugar composition (acyl chain composition and proportions of acylglucose and acylsucrose) (Shapiro et al., 1994). Our working hypothesis was that the differences in acylsugar amount and composition among different *S. pennellii* accessions are caused by one of the following factors: (i) differential expression of genes encoding the enzymes (and their regulatory factors) in the acylsugar biosynthetic pathway; and (ii) differences in transcript sequence and structure e.g., splice-isoforms, alternative start sites and/or alternative polyadenylation sites. RNA-Seq analysis has the potential to reveal these genetic differences. It has been demonstrated that differences in transcript levels of key enzymes is a major regulator of phenotype variation (glucosinolate accumulation in *Arabidopsis thaliana*) (Wentzell et al., 2007); therefore RNA-Seq tools were used to compare transcriptomes of low- and high-acylsugar-producing accessions of *S. pennellii* and identify genes that are differentially expressed. Four low-acylsugar-

producing accessions (LA1911, LA1912, LA1920, and LA1926), one medium-acylsugar-producing accession (LA1302), and three high-acylsugar-producing accessions of *S. pennellii* (LA0716, LA1941, and LA1946) were grown in greenhouse conditions. Leaf tissues from plants at an early-mid stage of development (~10-week old) were used for the extraction of both acylsugar and RNA from these accessions. After lipid extraction and purification, the amounts of secreted acylsugars were measured from each of the low-, medium-, and high-acylsugar-producing accessions to verify that these accessions are indeed low, medium, and high accessions (Table 2). RNA samples were extracted from leaf tissues of three or four individual plants (biological replicates) from these accessions. After performing DNase treatment and enrichment of mRNA samples, cDNA libraries were used to prepare Illumina Tru-Seq libraries. Sequencing was performed using the Illumina HiSeq 2500 platform (paired-end sequencing, 2x125 bp) at the Texas A&M AgriLife Genomics and Bioinformatics Services center. Illumina barcodes were used to distinguish different accessions and their biological replicates. Processed high-quality reads were mapped to the *S. pennellii* genome (Bolger et al., 2014a), and differentially expressed genes were identified using the edgeR program (Robinson et al., 2010; Chen et al., 2014; Zhou et al., 2014). Expression levels of interesting candidate genes were measured in the medium-acylsugar-producing accession LA1302 to check if expression levels of candidate genes and amounts of secreted acylsugars are positively correlated. Experimental methods and results for Specific aims 2-5 are discussed in detail in the following section of this dissertation.

Table 2. Amount of total acylsugars produced by different accessions of *S. pennellii*.

Accession	% Leaf Dry Weight	Standard Deviation
LA1911	0.45	0.09
LA1912	0.69	0.13
LA1920	0.55	0.23
LA1926	0.65	0.17
LA1302	5.62	0.96
LA1941	12.21	0.63
LA1946	12.57	0.35
LA0716	13.92	0.43

Amounts are represented as mean percentage of leaf dry weight (n=8).

1.9.2.2. Impact

Out of 44,966 annotated protein-coding genes in *S. pennellii* genome, 1087 were identified as differentially expressed between low- and high-acylsugar-producing accessions. Known and many high-confidence candidate genes showed intermediate expression levels in the medium-acylsugar-producing accession LA1302. Genes with putative functions in plant defense were also differentially expressed between low- and high-acylsugar-producing accessions; however, these putative defense genes did not show intermediate expression levels in LA1302.

1.9.3. Specific Aim#3: Differential Gene Expression Analysis to Filter Aim2 Candidate Genes That Respond to Imazapyr (Biochemical Inhibitor of Acylsugar Biosynthesis)

1.9.3.1. Background Information and Approach

Biosynthesis of specialized metabolites is tightly regulated (Kroymann, 2011; Owen et al., 2017). Production of plant defense compounds is often induced in response to pathogen and/or elicitor treatment, and this information has been exploited to identify specialized metabolic pathway genes (Giddings et al., 2011; Geu-Flores et al., 2012; Lau and Sattely, 2015). A different chemical genetics approach was taken to identify acylsugar metabolic genes. It has been demonstrated that inhibition of BCAA biosynthesis significantly inhibits acylsugar production in solanaceous plants (Kandra and Wagner, 1990; Walters and Steffens, 1990). We used a biochemical inhibitor of the enzyme acetolactate synthase (first common step in the biosynthesis of BCAAs), and identified genes that respond to the inhibitor treatment.

The sulfonylurea herbicides, such as chlorsulfuron and sulfometuron methyl, are highly potent commercial herbicides that can inhibit BCAA biosynthesis at nM levels (Scheel and Casida, 1985; Garcia et al., 2017). In *S. pennellii*, both of these two herbicides were administered as 10 µM solutions to block BCAA biosynthesis and subsequently inhibit acylsugar production (Walters and Steffens, 1990). Treatment with 1 µM chlorsulfuron in *Nicotiana tabacum* significantly reduced acylsugar production (Kandra and Wagner, 1990). On the other hand, the imidazolinone herbicides, such as

imazapyr (arsenal), are less powerful herbicides with K_i values in the µM ranges (Scheel and Casida, 1985; Garcia et al., 2017). However, due to solubility issues, water-soluble imazapyr was used in this study at two different concentrations (0.1 mM and 1 mM).

1.9.3.2. Impact

Known and high-confidence candidate genes exhibited concentration-dependent repression in response to imazapyr treatment. On the other hand, differentially expressed defense genes did not show such expression pattern. Using this information, the number of acylsugar candidate genes was reduced from 1087 to 171.

1.9.4. Specific Aim#4: Gene Co-Expression Network Analysis to Identify Acylsugar Metabolic Gene Network Using Gene Expression Profiles from Aim2 and Aim3 Experiments

1.9.4.1. Background Information and Approach

A gene network analysis is performed using gene expression data to identify interactions among genes. In such analysis, nodes and edges represent genes and their interactions, respectively. Usually three types of gene network analysis are performed for biological data: gene co-expression network analysis, association network analysis, and gene regulatory network analysis (Li et al., 2015; Sinha et al., 2016). In gene co-expression network analysis, co-expressed genes, whose expression profiles are strongly correlated, are connected by edges and clustered into modules. Connected genes have

direct or indirect interactions (no directionality). Such analysis can be useful for candidate gene prioritization (Li et al., 2015). Co-expressed genes are usually involved in the same biological process, such as biosynthesis of a specialized metabolite, and can be regulated by one or more common regulatory factors (transcription factors). In association networks, genes have direct interactions, but no directionality is provided. A gene co-expression network can be transformed into a gene regulatory network when edges have direct interactions and edges acquire directionality. Among these three types of gene interaction analysis methods, co-expression network analysis is the most versatile, and is particularly useful in plants where gene functions for most of the genes are unknown and information on regulatory interactions are unavailable (Sinha et al., 2016).

Normalized gene expression data, such as FPKM values, are used to determine pairwise correlation coefficients for expressed genes. Pearson correlation coefficient, which determines linear relationships between two variables, has been widely used to determine pairwise correlation coefficients in gene network analysis; however, Spearman's rank correlation coefficient is less sensitive to strong outliers and gives more robust results in case of non-linear relationships (Li et al., 2015; Sinha et al., 2016). Weighted gene correlation (or co-expression) network analysis (WGCNA) is a widely used biological gene network analysis program (R programming language software) (Langfelder and Horvath, 2008). WGCNA provides a weighted network in which edges are given a weight value. Hard-thresholding and dichotomizing (binary) methods lose useful biological information, whereas WGCNA preserves all the correlation

information among genes by using a soft thresholding method. In this approach, a network is represented by an adjacency matrix that reports the connection strength between gene pairs. Gene expression profile of *S. pennellii* genes in 38 samples were used for WGCNA analysis using Spearman's rank correlation coefficient values.

1.9.4.2. Impact

One co-expressed module with 182 genes was identified, which contained known and high-confidence candidate genes. 57 of these 182 genes responded to the imazapyr treatment at both concentrations. A co-expressed acylsugar gene network was proposed, which showed high connection strengths among genes. WGCNA also identified a defense gene module that contained genes involved in plant defense signaling pathways.

1.9.5. Specific Aim#5: Functional Validation of Two Candidate Genes Using Virus-Induced Gene Silencing Methods

1.9.5.1. Background Information and Approach

Functional genomics and transcriptomics provide a complete understanding of gene functions in a metabolic pathway or a physiological process. To achieve this, stable mutant lines can be used. However, mutant lines are available for only a handful of plant species. Gene overexpression or knockdown by RNAi are also used for elucidating gene functions, but these techniques are slow and can only be applied to plants for which well-known stable transformation protocols are available. Virus-induced gene silencing

(VIGS) has been widely used as a functional genomics tool in plants due to several advantages, such as fast gene silencing without the need for stable plant transformation, and due to the fact that VIGS can be used to silence multiple gene family members. Additionally, VIGS can be particularly useful in sexually propagated plants where gene mutations or knockdowns are embryo-lethal and/or can cause severe deformity in developing plants (Senthil-Kumar and Mysore, 2011).

VIGS achieves target gene(s) silencing by exploiting natural defense mechanisms of plants (Lu et al., 2003). It involves following three steps: engineering recombinant viruses that carry full or partial sequence of target gene(s), infecting the host plant, and achieving target gene silencing as a defense response of the host plant to viral infection. Plants protect themselves against viral infections by targeting viral nucleic acids for degradation via the RNAi pathway. VIGS is initiated when recombinant viral dsRNA containing target gene sequence is processed into small interfering RNA by dicer-like proteins, and RNAi is triggered that can silence plant target gene(s). This process deceives plants into identifying its own endogenous RNA as viral RNA, and subsequently post-transcriptional gene silencing is activated (Liu and Page, 2008; Senthil-Kumar and Mysore, 2011).

VIGS methods also have some challenges and limitations that need to be addressed when designing VIGS-mediated gene silencing experiments (Senthil-Kumar and Mysore, 2011). Availability of appropriate viral vectors and their delivery methods need to be checked. Sometimes viral infections can cause unwanted symptoms in host plants, which can lead to misinterpretation of actual gene silencing effects. When

designing VIGS constructs, it should be ensured that there is no possible off-target effects, or at least off-target possibility should be minimized.

Several VIGS vectors are available for gene silencing, and tobacco rattle virus (TRV)-based VIGS vectors have been used in solanaceous plants, such as *N. benthamiana* and *S. lycopersicum* (Senthil-Kumar and Mysore, 2011). TRV is a bipartite plant virus, and its genome consists of two linear positive sense ssRNA molecules (Bachan and Dinesh-Kumar, 2012; Senthil-Kumar and Mysore, 2014). RNA1 is about 6.8 kb and RNA2 ranges from 1.8 kb to about 4.5 kb in size (different sizes in different isolates). RNA1 is essential for viral movement, and it has genes encoding RNA-dependent RNA polymerase, a movement protein and a cysteine-rich protein. RNA2 contains genes encoding the coat protein and nonstructural proteins that are not essential for plant infection. These non-essential genes from the RNA2 have been replaced with multiple cloning sites to produce TRV-based VIGS vectors (Senthil-Kumar and Mysore, 2014). TRV1- and TRV2-based binary vectors are cloned in *Agrobacterium tumefaciens*, and gene silencing can be achieved in 3-4 weeks (Dong et al., 2007; Senthil-Kumar and Mysore, 2014).

The *phytoene desaturase* gene (*PDS*) is commonly used as a marker in VIGS experiments (Liu and Page, 2008). PDS is involved in carotenoid biosynthesis. This enzyme catalyzes the conversion of colorless 15-*cis*-phytoene into bright red-colored lycopene. Gene silencing of the *PDS* leads to photobleaching (white leaf tissues). The efficiency of VIGS silencing is determined through the *PDS* silencing and subsequent photobleaching. Several factors, such as target gene fragment size, position of the target

region with respect to the full-length transcript, and presence of homopolymeric regions, can influence the efficiency of VIGS (Liu and Page, 2008). For example, in *N. benthamiana*, target fragments longer than 1.5 kb give little or no silencing, possibly because viral replication is affected by the large size of foreign inserts. On the other hand, inserts smaller than 100 bp result in inefficient silencing. Target sites closer to the extreme ends of mRNA (5' UTR and 3' UTR) should be avoided, if possible, and homopolymeric regions significantly reduce VIGS efficiency (Liu and Page, 2008). In this work, TRV-based vectors were used for gene silencing of two acylsugar candidate genes. The *phytoene desaturase* gene from *S. pennellii* was used as a marker to measure the efficiency of VIGS.

1.9.5.2. Impact

VIGS of two candidate genes for straight-chain fatty acid biosynthesis confirmed their role in acylsugar metabolism, and provided functional validation. This will strongly encourage functional validation of remaining high-confidence candidate genes using VIGS.

2. CANDIDATE GENE NETWORKS FOR ACYLSUGAR METABOLISM AND PLANT DEFENSE IN WILD TOMATO *Solanum pennellii**

2.1. Overview

Many solanaceous plants secrete acylsugars, which are branched-chain and straight-chain fatty acids esterified to glucose or sucrose. These compounds have important roles in plant defense and potential commercial applications. However, several acylsugar metabolic genes remain unidentified, and little is known about regulation of this pathway. Comparative transcriptomics between low- and high-acylsugar-producing accessions of *Solanum pennellii* revealed that expression levels of known and novel candidate genes (putatively encoding beta-ketoacyl-(acyl-carrier-protein) synthases, peroxisomal acyl-activating enzymes, ABC transporters, and central carbon metabolic proteins) were positively correlated with acylsugar accumulation, except two genes previously reported to be involved in acylglucose biosynthesis. Genes putatively encoding oxylipin metabolic proteins, subtilisin-like proteases, and other antimicrobial defense proteins were upregulated in low-acylsugar-producing accessions. Transcriptome analysis after biochemical inhibition of biosynthesis of branched-chain amino acids (precursors to branched-chain fatty acids) by imazapyr showed concentration-dependent downregulation of known and most acylsugar candidate genes,

*Part of the data reported in this section is reprinted with permission from "Acylsugar and plant defense gene networks" by Sabyasachi Mandal, Wangming Ji, and Thomas D. McKnight, 2019. *The Plant Cell* Oct 2019, tpc.00552.2019; DOI: 10.1105/tpc.19.00552 (www.plantcell.org). Copyright 2019 by American Society of Plant Biologists.

but not defense genes. Weighted gene correlation network analysis identified separate co-expressed gene networks for acylsugar metabolism (including six transcription factor genes and flavonoid metabolic genes) and plant defense (including genes putatively encoding NB-ARC and leucine-rich repeat sequences, protein kinases and defense signaling proteins, and previously mentioned defense proteins). Additionally, virus-induced gene silencing of two trichome preferentially expressed candidate genes for straight-chain fatty acid biosynthesis confirmed their role in acylsugar metabolism.

2.2. Introduction

Plants synthesize a wide variety of specialized metabolites that provide selective advantages in specific environmental conditions. Different classes of specialized metabolites are found in specific taxonomic groups, and many of them are valuable phytochemicals. Acylsugars are nonvolatile and viscous specialized metabolites secreted through trichomes of solanaceous plants (Fobes et al., 1985; Kroumova et al., 2016; Moghe et al., 2017), and their roles in plant defense have been studied extensively. For example, acylsugars from *Solanum pennellii* act as feeding and oviposition deterrents for insect pests and exert toxic effects on different herbivores (Hawthorne et al., 1992; Juvik et al., 1994; Liedl et al., 1995). Acylsugars from other genera also have important roles in providing protection against herbivores and plant pathogens (Chortyk et al., 1997; Hare, 2005; Luu et al., 2017). Additionally, amphipathic acylsugar molecules are believed to reduce the surface tension of adsorbed dew, thereby providing an additional source of water for the plants (Fobes et al., 1985). Acylsugars also have potential

applications as pesticides (Puterka et al., 2003), food additives (Hill and Rhode, 1999), cosmetic and personal care products (Hill and Rhode, 1999), antibiotics (Chortyk et al., 1993), and anti-inflammatory compounds (Perez-Castorena et al., 2010). Due to these important biological roles and potential commercial applications of acylsugars, a detailed understanding of the genetic network involved in acylsugar metabolism is required for successful crop breeding programs and metabolic engineering of acylsugar production.

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

Biosynthesis of acylsugars can be broadly divided into two phases: 1) synthesis of fatty acyl chains, and 2) esterification of these acyl molecules to glucose or sucrose (Figure 3A). Three acylsugar acyltransferases (ASATs) are involved in phase 2 of acylsucrose biosynthesis (Schilmiller et al., 2015; Fan et al., 2016; Fan et al., 2017).

During the preparation of this dissertation, an invertase (acylsucrose fructofuranosidase 1; ASFF1) was reported to be capable of producing acylglucose from acylsucrose (Leong et al., 2019). It should be noted that previous biochemical studies reported a UDP-glucose:fatty acid glucosyltransferase (UDP-Glc:FA GT) (Ghangas and Steffens, 1993; Kuai et al., 1997) and a serine carboxypeptidase-like glucose acyltransferase (SCPL GAT) (Li et al., 1999; Li and Steffens, 2000) that initiate acylglucose biosynthesis in *S. pennellii* (Figure 3A). However, to our knowledge, no genetic evidence is available to support the involvement of these two genes in the pathway.

Acylsugars are exuded through trichomes, predominantly by type IV trichomes in *S. pennellii* (Fobes et al., 1985; Slocombe et al., 2008), and acylsucrose biosynthetic genes are expressed in trichome tip cells (Schilmiller et al., 2012; Schilmiller et al., 2015; Fan et al., 2016). However, there are no reports of acylsugar biosynthesis in isolated *S. pennellii* trichomes. In fact, analysis of periclinal chimeras indicates that acylglucose biosynthesis in *S. pennellii* is not trichome specific (Kuai et al., 1997). Therefore, we used leaf samples for comparative transcriptomics studies reported here. Although this approach underrepresents actual expression differences of trichome-specific genes, it ensures that a complete snapshot of a candidate gene network could be captured.

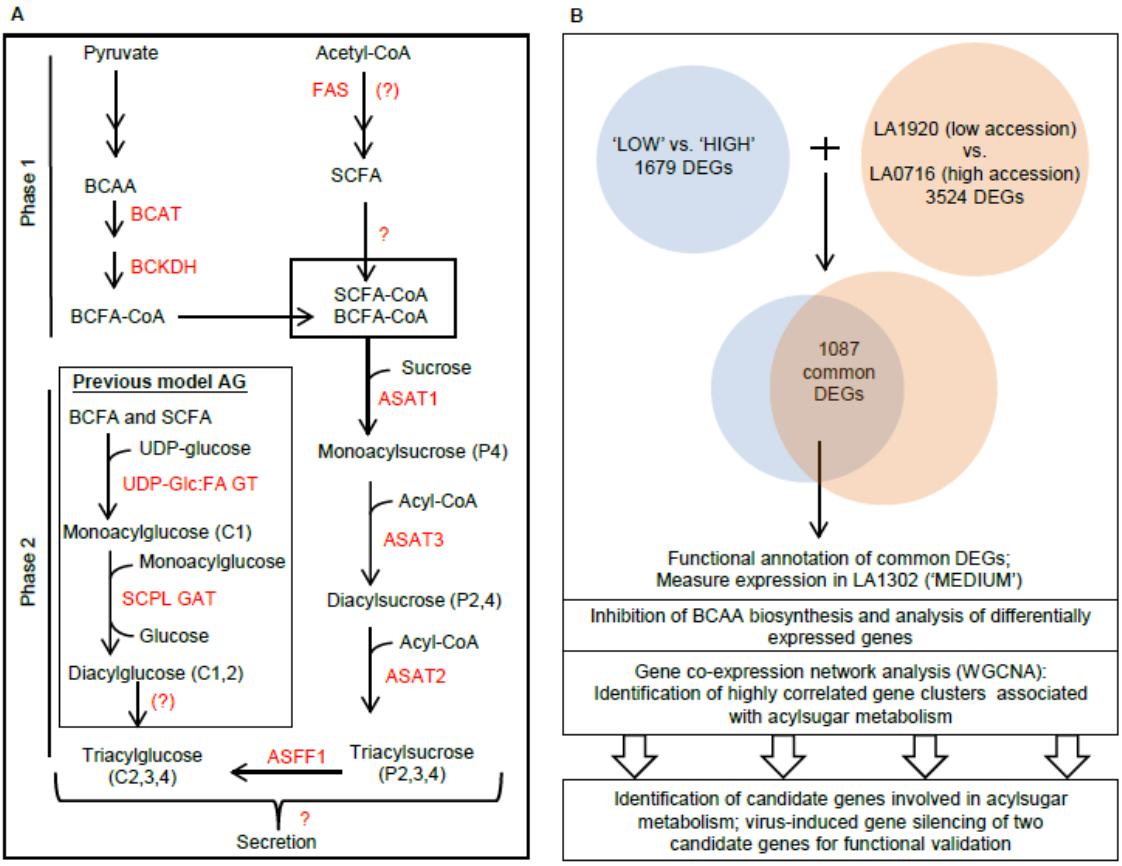

Figure 3. Current model of acylsugar production in *Solanum pennellii* and workflow to identify candidate genes for acylsugar metabolism. (A) Previous model of acylglucose (AG) biosynthesis is shown within a box. Enzymes are highlighted in red. Unidentified enzymes and transporters are marked with '?'. 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'). WGCNA indicates weighted gene correlation network analysis, respectively.

Different accessions of *S. pennellii* produce different amounts of total acylsugars (Shapiro et al., 1994). Certain accessions of *S. pennellii*, for example LA0716, can produce acylsugars up to 20% of its leaf dry weight (Fobes et al., 1985), whereas in many low-acylsugar-producing accessions, acylsugars make up less than 1% of leaf dry weight (Shapiro et al., 1994). Here, we exploited this natural genetic variation among *S. pennellii* accessions to identify candidate biosynthetic and regulatory genes for this trait. Comparative transcriptomics following biochemical inhibition of BCAA biosynthesis further refined the list of candidate genes and shed light on possible regulatory mechanisms of acylsugar production. Functional analysis through virus-induced gene silencing (VIGS) of two candidate genes for straight-chain fatty acid biosynthesis corroborated their involvement in acylsugar metabolism, and validated our approach to identifying new genes involved in this pathway. Additionally, gene co-expression network analysis revealed genetic networks involved in acylsugar metabolism and plant defense pathways.

2.3. Results

2.3.1. Comparative Transcriptomics between Low- and High-Acylsugar-Producing Accessions

We performed differential gene expression analysis between a group of three low-acylsugar-producing accessions (LA1911, LA1912, and LA1926; collectively referred to as ‘LOW’) and a group of three high-acylsugar-producing accessions

(LA1941, LA1946, and LA0716; collectively referred to as 'HIGH') of *S. pennellii* (Shapiro et al., 1994). Three individual plants for each accession, with an average of >30 million 125-bp paired-end reads for each plant, were used for differential gene expression analysis ('LOW', n=9; 'HIGH', n=9). 91-97% of the processed reads mapped to the *S. pennellii* genome, with 5-7% of the mapped reads having multiple-hits (Table 3), which were removed from the mapped-read counting process and differential gene expression testing by edgeR. 19,379 genes passed our filtering criteria for minimum expression levels, and we obtained 1679 differentially expressed genes (DEGs). Of the 1679 DEGs, 931 were upregulated and 748 were downregulated in the 'HIGH' group. The list of DEGs and their annotation is available as Supplemental Dataset 1 at http://www.plantcell.org/content/early/2019/10/18/tpc.19.00552/tab-figures-data.

Table 3. RNA-Seq read-mapping results (TopHat2 mapping of filtered reads).

Accession_replicate	Input reads (paired-end)	Left reads mapped	% Left reads mapped	Right reads mapped	% Right reads mapped	Aligned pairs	% Multiple alignments
LA1911_1	28076632 x2	25670656	91.43	25460139	90.68	24479209	5.73
LA1911_2	25897024 x2	23979058	92.59	23833753	92.03	22916745	5.92
LA1911_3	23381377 x2	21606735	92.41	21468199	91.82	20477277	6.25
LA1912_1	26513055 x2	24446188	92.20	24223469	91.36	23238773	5.96
LA1912_2	27459168 x2	25464977	92.74	25350740	92.32	24300336	6.31
LA1912_3	27197768 x2	25230181	92.77	25091449	92.26	24040708	5.87
LA1926_1	24579536 x2	22652595	92.16	22408034	91.17	21458492	6.07
LA1926_2	25068349 x2	23131746	92.27	22993018	91.72	22011845	6.35

Table 3 continued.

Accession_replicate	Input reads (paired-end)	Left reads mapped	% Left reads mapped	Right reads mapped	% Right reads mapped	Aligned pairs	% Multiple alignments
LA1926_3	24724627 x2	22833110	92.35	22734307	91.95	21738682	5.55
LA1941_1	12288246 x2	11502460	93.61	11432671	93.04	11179476	4.83
LA1941_2	26080846 x2	24793307	95.06	24642137	94.48	23941692	6.34
LA1941_3	26040582 x2	24769968	95.12	24691874	94.82	23981049	6.36
LA1946_1	24173944 x2	22771758	94.20	22569746	93.36	21959695	5.86
LA1946_2	31780978 x2	30153635	94.88	29916301	94.13	29036842	6.95
LA1946_3	29730066 x2	28193764	94.83	28075357	94.43	27248926	6.13
LA0716_1	30522535 x2	29677629	97.23	29484523	96.60	29001962	5.96
LA0716_2	33592217 x2	32412058	96.49	32213141	95.89	31522686	6.95
LA0716_3	28833828 x2	28017182	97.17	27920559	96.83	27454445	5.94
LA1302_1	22253612 x2	20010595	89.92	19866310	89.27	19017544	5.82
LA1302_2	26994304 x2	24817574	91.94	24703973	91.52	23593535	7.20
LA1302_3	26601942 x2	24436183	91.86	24316523	91.41	23197918	6.39
LA1920_1	25852990 x2	23916352	92.51	23823834	92.15	22843819	11.29
LA1920_2	27558000 x2	25525014	92.62	25367999	92.05	24316564	6.56
LA1920_3	41277779 x2	38095756	92.29	38002335	92.06	36328058	6.24
LA1920_4	52957983 x2	48942356	92.42	48714958	91.99	46652941	6.43
LA0716_1	28097123 x2	27323686	97.25	27106013	96.47	26656972	5.74
LA0716_2	26072342 x2	24860684	95.35	24722176	94.82	24324913	7.19
LA0716_3	48846961 x2	47425319	97.09	47182276	96.59	46348049	5.60
LA0716_4	54976879 x2	53381600	97.10	53172897	96.72	52208391	5.86

We also compared transcriptomes between another low-acylsugar-producing accession, LA1920, and the high-acylsugar-producing accession LA0716, with each accession having four biological replicates and slightly higher sequencing coverage, as

an independent examination of differential gene expression. In this analysis, 19,967 genes passed our filtering criteria for minimum expression levels, and we identified 3524 DEGs (1465 upregulated and 2059 downregulated genes in LA0716). The higher number of DEGs in this two-accession comparison likely includes many accession-specific DEGs that would not be apparent when multiple 'LOW' accessions are compared against multiple 'HIGH' accessions.

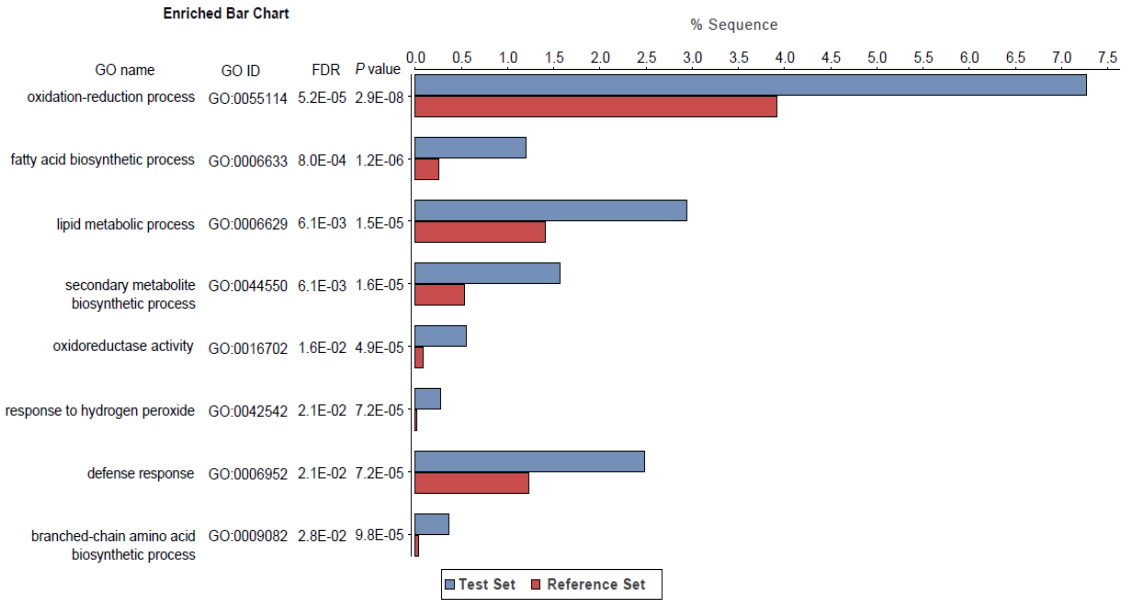

Figure 4. Enriched gene ontology (GO) terms associated with the 1087 common DEGs. Test set and reference set indicate 1087 DEGs and *S. pennellii* annotated sequences, respectively. Enrichment analysis was performed using Blast2GO software (Fischer's exact test). Only significant terms are shown (false discovery rate, FDR < 0.05).

To identify candidate genes most likely to be involved in acylsugar metabolism, we compared 'LOW' vs. 'HIGH' DEGs with LA1920 vs. LA0716 DEGs (Figure 3B).

1087 DEGs were common to both comparisons (586 upregulated and 501 downregulated genes in high-acylsugar-producing accessions), and these 1087 DEGs corresponded to 1198 transcript sequences according to *S. pennellii* gene models (Bolger et al., 2014a). This approach was designed to eliminate accession-specific DEGs that were presumably not critical to acylsugar metabolism. For instance, *Sopen00g011600* (aspartyl protease) and *Sopen02g031180* (hypothetical protein) had 1100- and 390-fold higher expression in the 'HIGH' group (FDR= 3.09E-08 and 9.61E-07, respectively), but they were excluded from the list of common DEGs because they were expressed in LA1941 and LA1946 only, but not in LA0716. Gene Ontology (GO) terms associated with the annotated DEGs were extracted, and enrichment analysis revealed significant overrepresentation of GO terms such as 'secondary metabolite biosynthetic process' (GO:0044550), 'fatty acid biosynthetic process' (GO:0006633), 'branched-chain amino acid biosynthetic process' (GO:0009082), and 'defense response' (GO:0006952) in the list of common DEGs (Figure 4).

We also examined the leaf transcriptome of accession LA1302 (three biological replicates), which accumulates an intermediate level of acylsugars (referred to as 'MEDIUM' in this thesis) (Shapiro et al., 1994), to determinate whether expression levels of candidate genes were consistent with the amount of acylsugar production (Figure 3B). Multi-dimensional scaling plots showed a clear distinction in gene expression profiles among groups of accessions that accumulate different amounts of acylsugars (Figure 5A and 5B). Volcano plots showed high $\log_2$(fold-change) values associated with low false discovery rate (FDR) values, indicating robust differential

expression (Figure 5C and 5D). On the other hand, ten ‘housekeeping genes’ showed similar expression levels across all biological groups (Figure 6).

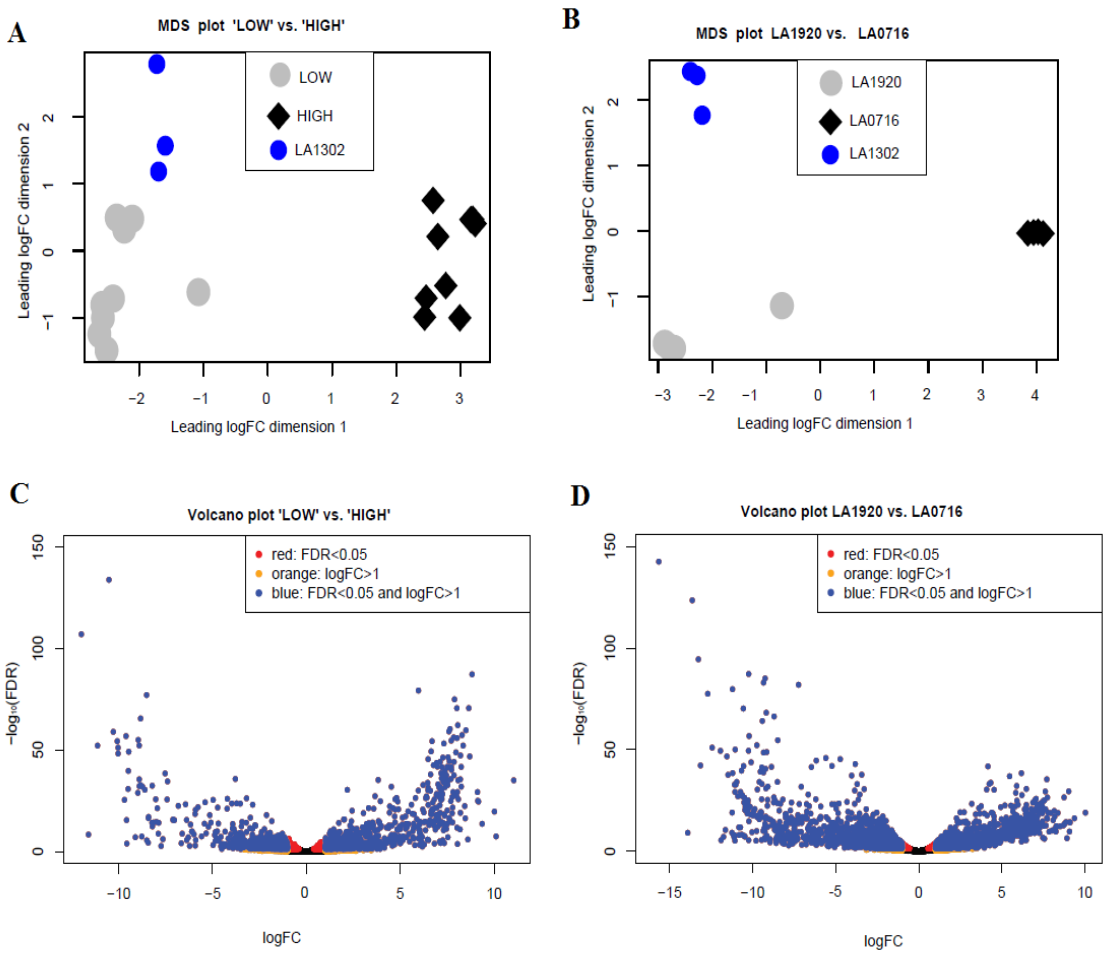

Figure 5. Differential gene expression analyses between low- and high-acylsugar-producing accessions. (A) and (B) Multi-dimensional scaling (MDS) plots showing ‘LOW’ vs. ‘HIGH’ comparison and LA1920 vs. LA0716 comparison, respectively. Three biological replicates of LA1302 were also included to show their gene expression profiles relative to low- and high-acylsugar-producing accessions. Distances were determined by edgeR based on expression profiles of top-500 genes differing in fold changes. (C) and (D) Volcano plots showing ‘LOW’ vs. ‘HIGH’ comparison and LA1920 vs. LA0716 comparison, respectively. Positive logFC values indicate higher expression levels in high-acylsugar-producing accessions. Abbreviations: FDR= false discovery rate; logFC= $\log_2$(fold-change).

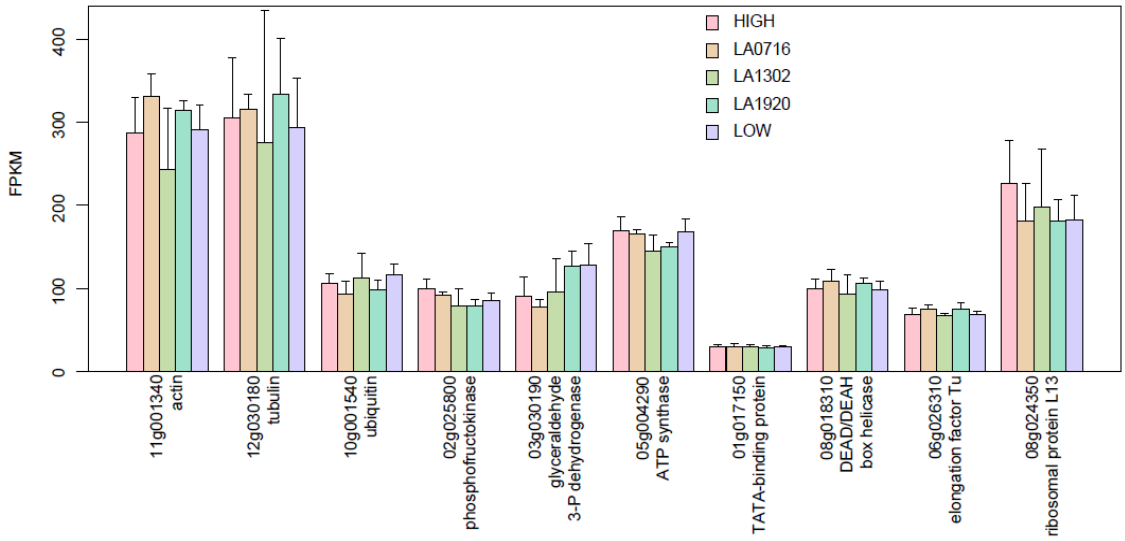

Figure 6. Expression levels of ten selected 'housekeeping genes' in different biological groups. Error bars represent 95% confidence interval (only upper limits are shown). Abbreviation: FPKM= fragments per kilobase of transcript per million mapped reads.

2.3.2. Genes Involved in Branched-Chain Amino Acid (BCAA) Metabolism Are Upregulated in High-Acylsugar-Producing Accessions

Isotope-feeding studies with branched-chain amino acids (BCAAs) and their keto derivatives showed that the BCAAs play an important role in the biosynthesis of acylsugar in *S. pennellii* (Walters and Steffens, 1990). Using inhibitors of BCAA biosynthetic enzyme, it was further concluded that BCAAs contribute to the branched-chain acyl moieties of acylsugars. Therefore, we examined genes involved in BCAA metabolism.

In plants, BCAAs (valine, isoleucine, and leucine) are synthesized in chloroplasts, and biosynthesis of valine and isoleucine share a common set of four enzymes (Binder, 2010; Galili et al., 2016). We found genes encoding these enzymes to

be up-regulated in high-acylsugar-producing accessions. The first three of these four enzymes are acetohydroxyacid synthase (AHAS, also known as acetolactate synthase ALS; EC number 2.2.1.6), acetohydroxyacid isomeroreductase (AHIR, also known as ketolacid reductoisomerase KARI; EC number 1.1.1.86), and dihydroxyacid dehydratase (DHAD; EC number 4.2.1.9). Three genes (*Sopen11g004560, Sopen07g027240,* and *Sopen05g032060*) putatively encoding these enzymes had higher expression levels (2.65-, 2.72-, and 3.39-fold, respectively; Figure 7A) in the 'HIGH' accessions. The final products of these three enzymatic steps are 3-methyl-2-oxobutanoate and 3-methyl-2-oxopentanoate, which are the keto derivatives of valine and isoleucine, respectively. Leucine is synthesized from 3-methyl-2-oxobutanoate, and the committed step in Leucine biosynthesis is catalyzed by the enzyme isopropylmalate synthase (IPMS; EC number 2.3.3.13) (de Kraker et al., 2007). Two genes encoding IPMS were up-regulated in the high-acylsugar-producing accessions. One of them (*Sopen08g005060*) had only 2-fold higher expression (FPKM of 52 and 105, in 'LOW' and 'HIGH' groups, respectively), whereas the other one (*Sopen08g005140*) had 32-fold higher expression (FPKM of 0.33 and 10.58, in 'LOW' and 'HIGH' groups, respectively). *Sopen08g005140* encodes a variant IPMS which is involved in specialized metabolism, and this enzyme affects acylsugar composition in *S. pennellii* and *S. lycopersicum* (Ning et al., 2015). Additionally, the tomato homolog of *Sopen08g005140* (*Solyc08g014230*) is expressed in the apical cells of type I/IV trichomes in *S. lycopersicum*; therefore, the expression difference of this gene between 'LOW' and 'HIGH' group could have been downplayed by the presence of underlying tissues during RNA-Seq library preparation.

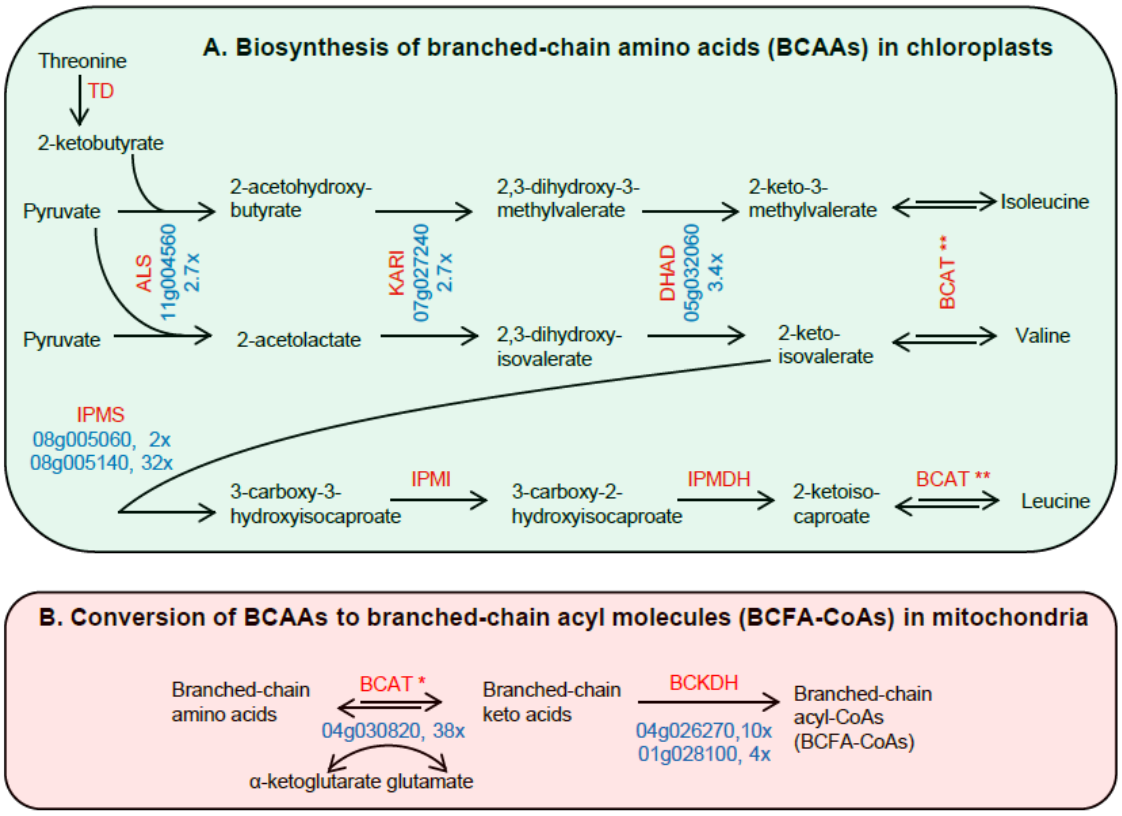

Figure 7. Generation of acylsugar branched-chain acyl groups. Enzymes are highlighted in red. Gene identifiers of differentially expressed genes and their fold-changes are highlighted in blue. 10x indicates 10-fold higher expression in the 'HIGH' group relative to the 'LOW' group. Mitochondrial and chloroplast BCATs are marked with one and two asterisks, respectively. Abbreviations: TD= threonine dehydratase; ALS= acetolactate synthase; KARI= ketol-acid reductoisomerase; DHAD= dihydroxy-acid dehydratase; IPMS= isopropylmalate synthase; IPMI= isopropylmalate isomerase; IPMDH= isopropylmalate dehydrogenase; BCAT= branched-chain aminotransferase; BCKDH= branched-chain keto acid dehydrogenase.

Conversion of BCAAs to branched-chain acyl molecules is initiated by a transamination reaction catalyzed by branched-chain aminotransferase (BCAT; EC number 2.6.1.42). One gene (*Sopen04g030820*) encoding BCAT had 38-fold higher expression (FPKM of 0.62 and 24.6, in 'LOW' and 'HIGH' groups, respectively). It

should be noted that BCATs also catalyze the last common step in the biosynthesis of all three BCAAs. In *S. lycopersicum*, seven genes encoding BCATs have been identified and characterized (Maloney et al., 2010; Kochevenko et al., 2012). Subcellular localization studies using green fluorescence protein and complementation studies with *Escherichia coli* auxotrophic cells showed that the mitochondrial BCAT-1 and BCAT-2 are involved primarily in BCAA to BCFA conversion, whereas chloroplast BCAT-3 and BCAT-4 function primarily in BCAA biosynthesis. BLAST results reveal that *Sopen04g030820* shows 87% similarity (98% coverage, e-value 0.0) with Sl-*BCAT2*, suggesting its mitochondrial location (Figure 7).

The second step in BCAA to BCFA-CoA conversion is catalyzed by the branched-chain keto acid dehydrogenase (BCKDH; EC number 1.2.4.4), which converts the branched-chain keto acids into acyl-CoA esters. This multienzyme complex shows structural homology to pyruvate dehydrogenase complex and α-ketoglutarate dehydrogenase complex, and is composed of three subunits: α-keto acid dehydrogenase/carboxylase E1 (composed of E1α and E1β; catalyzes the decarboxylation of α-keto acids), dihydrolipoyl acyltransferase E2 (forms HS-CoA esters), and dihydrolipoyl dehydrogenase E3 (oxidizes HS-CoA into acyl-CoA) (Mooney et al., 2002). In *Arabidopsis thaliana* primary metabolism, BCKDH catalyzes the committed step in BCAA catabolism, and this mitochondrial enzyme complex plays an important role in regulating amino acid homeostasis (Taylor et al., 2004; Peng et al., 2015). Two genes encoding E1 and E2 components of BCKDH (*Sopen04g026270* and *Sopen01g028100*, respectively) had higher expression levels in high-acylsugar-

producing accessions (Table 4). Virus induced gene silencing studies with E1β component of BCKDH complex showed that BCKDH plays an important role in the production of branched-chain acyl groups in *S. pennellii* and *Nicotiana benthamiana* (Slocombe et al., 2008). Higher expression levels of genes encoding components of BCKDH complex further supports the role of this enzyme complex in providing branched-chain acyl-CoA molecules for incorporation into acylsugars.

It should be noted that BCAAs can be utilized to support mitochondrial respiration also (catabolic steps converting BCFA-CoAs to citric acid cycle intermediates in plant primary metabolism). For example, acetyl-CoA can be generated in mitochondria by completely degrading leucine and isoleucine, in a way similar to β-oxidation (Binder, 2010). We did not find any DEG involved either in degradation of branched-chain acyl-CoA derivatives (BCFA-CoAs) of leucine and isoleucine, or in subsequent catabolic steps leading to acetyl-CoA formation. However, two DEGs (*Sopen07g023250* and *Sopen05g023470*) predicted to encode 3-hydroxyisobutyryl-CoA hydrolase, an enzyme involved in degradation of valine, had 27- and 2.5-fold higher expression, respectively in the 'HIGH' group (FDR= 1.40E-08 and 1.40E-02, respectively). Complete degradation of valine ultimately produces succinyl-CoA, a citric acid cycle intermediate produced from α-ketoglutarate. Mitochondrial BCATs, which catalyze the transamination reaction of BCAA to BCFA-CoA conversion, also utilize α-ketoglutarate to make glutamate (Figure 7B), and this consumption of α-ketoglutarate may be responsible for the anaplerotic response in the form of valine degradation.

Table 4. Differentially expressed genes involved in phase 1 of acylsugar biosynthesis.

Gene	'LOW' vs. 'HIGH'		LA1920 vs. LA0716		Annotation
	Log_2FC	FDR	Log_2FC	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
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

Log_2FC and FDR indicate log_2 (fold-change) and false discovery rate (*P* values adjusted for multiple-testing), respectively. Positive and negative log_2FC values indicate higher and lower expression levels, respectively in high-acylsugar-producing accessions.

2.3.3. Genes Putatively Encoding a KAS IV/ KAS II-Like Enzyme and Other FAS Components are Upregulated in High-Acylsugar-Producing Accessions

Sopen12g004230, predicted to encode a beta-ketoacyl-(acyl-carrier-protein) synthase II (KAS II)-like enzyme, had 39-fold higher expression in the 'HIGH' group (FDR= 4.40E-18). The SCFA profile of *S. pennellii* acylsugars (predominantly C10 and C12, with a trace amount of C8 in LA1302) is strikingly similar to that of the seed oil in many *Cuphea* species (Shapiro et al., 1994; Dehesh et al., 1998; Slabaugh et al., 1998; Schutt et al., 2002). A specialized version of KAS II, referred to as KAS IV, is an important determinant of chain length in *Cuphea* medium-chain (C8-C12) fatty acids (Dehesh et al., 1998; Slabaugh et al., 1998; Schutt et al., 2002). We compared amino acid sequences of five KAS IV/ KAS II-like enzymes from four *Cuphea* species with the Sopen12g004230 sequence, but no similarity was found. However, sequence analysis of the adjacent Sopen12g004240 coding region (1785-aa; annotated as reverse transcriptase) revealed a KAS domain at the C-terminal end (1374-1784), and BLAST analysis of *Cuphea* KAS IV/ KAS II-like enzymes showed high sequence similarity with Sopen12g004240 (77%-83% identity, 87%-92% similarity, 82%-90% coverage, e-value 0.0; Table 5). Apparently, a transposon inserted into an intron of *Sopen12g004230-Sopen12g004240*, which led to misannotation of this locus as two separate genes. We confirmed that this locus produces a single transcript with RT-PCR analysis (Figure 8), and Sanger sequencing confirmed that this single transcript can produce a functional KAS protein. Blauth et al. (1999) used an intraspecific F_2 population derived from the cross between accessions LA0716 and LA1912, to identify a quantitative trait locus on

the upper arm of chromosome 12 that is associated with C10 SCFA production (Blauth et al., 1999). The locus containing *Sopen12g004230-Sopen12g004240* is in this region.

DEGs putatively encoding other components of the FAS complex were also upregulated in high-acylsugar-producing accessions (Figure 9). Most of the BCAA/BCFA and SCFA metabolic DEGs showed intermediate levels of expression in the 'MEDIUM' accession LA1302 (Figure 10).

Table 5. Sequence similarity between *Cuphea* KAS IV/ KAS II-like enzymes and Sopen12g004230-Sopen12g004240.

			Sopen12g001230				**Sopen12g001240**			
GenBank accession	**Cuphea species**	**Enzyme (length)**	**% Coverage**	**% Identity**	**% Similarity**	**e-value**	**% Coverage**	**% Identity**	**% Similarity**	**e-value**
AF060518	*C. pulcherrima*	KAS IV (545-aa)	12	24	43	1.0	82	83	92	0.0
AF060519	*C. hookeriana*	KAS IV (533-aa)	3	50	80	0.3	90	77	87	0.0
U67317	*C. wrightii*	KAS II (540-aa)	13	47	58	4.1	86	82	90	0.0
U67316	*C. wrightii*	KAS II (517-aa)	11	22	43	3.0	85	82	90	0.0
AJ344250	*C. lanceolata*	KAS IV (538-aa)	7	36	54	4.7	83	81	89	0.0

Figure 8. Structure of the locus containing Sopen12g004230-Sopen12g004240. Integrated Genome Viewer (Thorvaldsdottir et al., 2013) was used to visualize RNA-Seq reads mapped to *S. pennellii* chromosome 12 locus containing *Sopen12g004230-Sopen12g004240*.

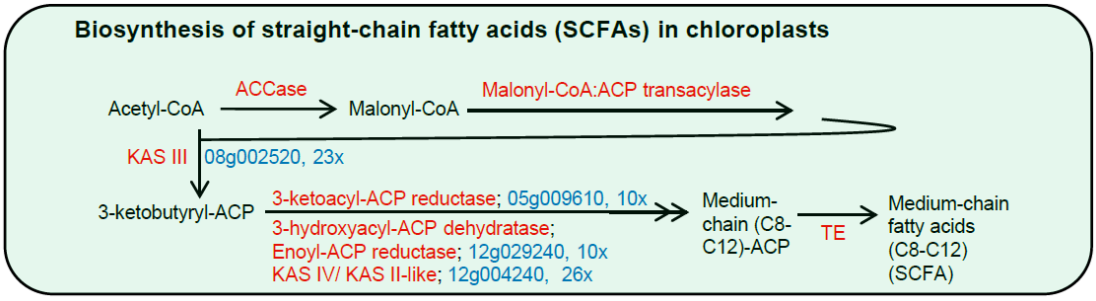

Figure 9. Generation of acylsugar straight-chain acyl groups in chloroplasts. Enzymes are highlighted in red. Gene identifiers of differentially expressed genes and their fold-changes are highlighted in blue. 10x indicates 10-fold higher expression in the 'HIGH' group relative to the 'LOW' group. Abbreviations: ACCase= acetyl-CoA carboxylase; KAS= beta-ketoacyl-ACP synthase; TE= thioesterase; ACP= acyl carrier protein.

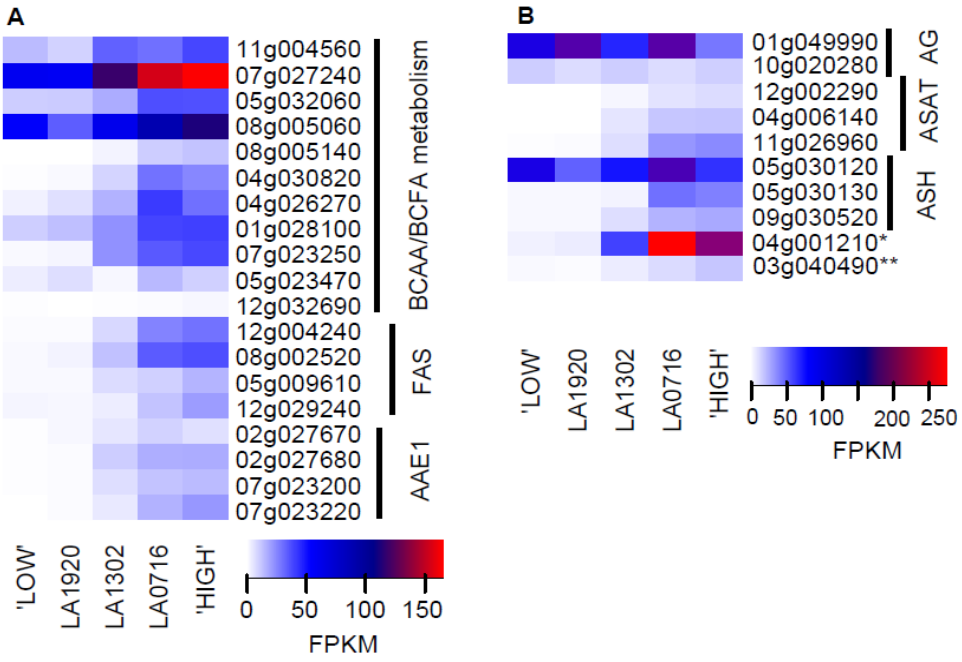

Figure 10. Heatmaps showing expression levels of phase 1- (A) and phase 2-related genes (B). *Sopen04g001210* (marked with an asterisk) was identified as a carboxylesterase gene related to ASH (Schilmiller et al., 2016). *Sopen03g040490* (marked with double-asterisk) is the invertase gene. Abbreviations: FPKM= fragments per kilobase of transcript per million mapped reads; BCAA= branched-chain amino acid; BCFA= branched-chain fatty acid; FAS= fatty acid synthase components; AAE1= acyl-activating enzyme 1; AG= acylglucose phase 2 (previous model); ASAT= acylsucrose acyltransferase; ASH= acylsugar hydrolase.

2.3.4. Four Genes Putatively Encoding Acyl-Activating Enzymes Show Strong Positive Correlation with FAS Components

Acyl-activating enzymes (AAEs) are ATP/AMP-binding proteins that activate different carboxylic acids by forming fatty acyl-CoA molecules from free fatty acids, ATP and CoA (Shockey et al., 2003). We found four DEGs (*Sopen02g027670*, *Sopen02g027680*, *Sopen07g023200*, and *Sopen07g023220*) putatively encoding acyl-activating enzyme 1 (AAE1) that had 8- to 56-fold higher expression in the 'HIGH' group (FDR= 2.80E-07 to 3.80E-27; Table 4). These genes are located on chromosome 2 and chromosome 7 as members of similar gene clusters. Biochemical and phylogenetic studies in *Arabidopsis thaliana* showed that AAE7 and AAE11, which belong to the same phylogenetic clade as AAE1, preferentially use short- and medium-chain fatty acids as their substrates (Shockey et al., 2003). This suggests that DEGs putatively encoding AAE1 in *S. pennellii* may also exhibit selective activity against C4-C12 fatty acids. Additionally, correlation analysis (Spearman's rank correlation coefficient, SRCC; denoted by ρ) using gene expression profiles in our 29 samples revealed that these four AAE1 genes exhibit strong positive correlation with FAS components (ρ= 0.87 to 0.98; P = 1.07E-9 to 4.44E-16; Figure 11).

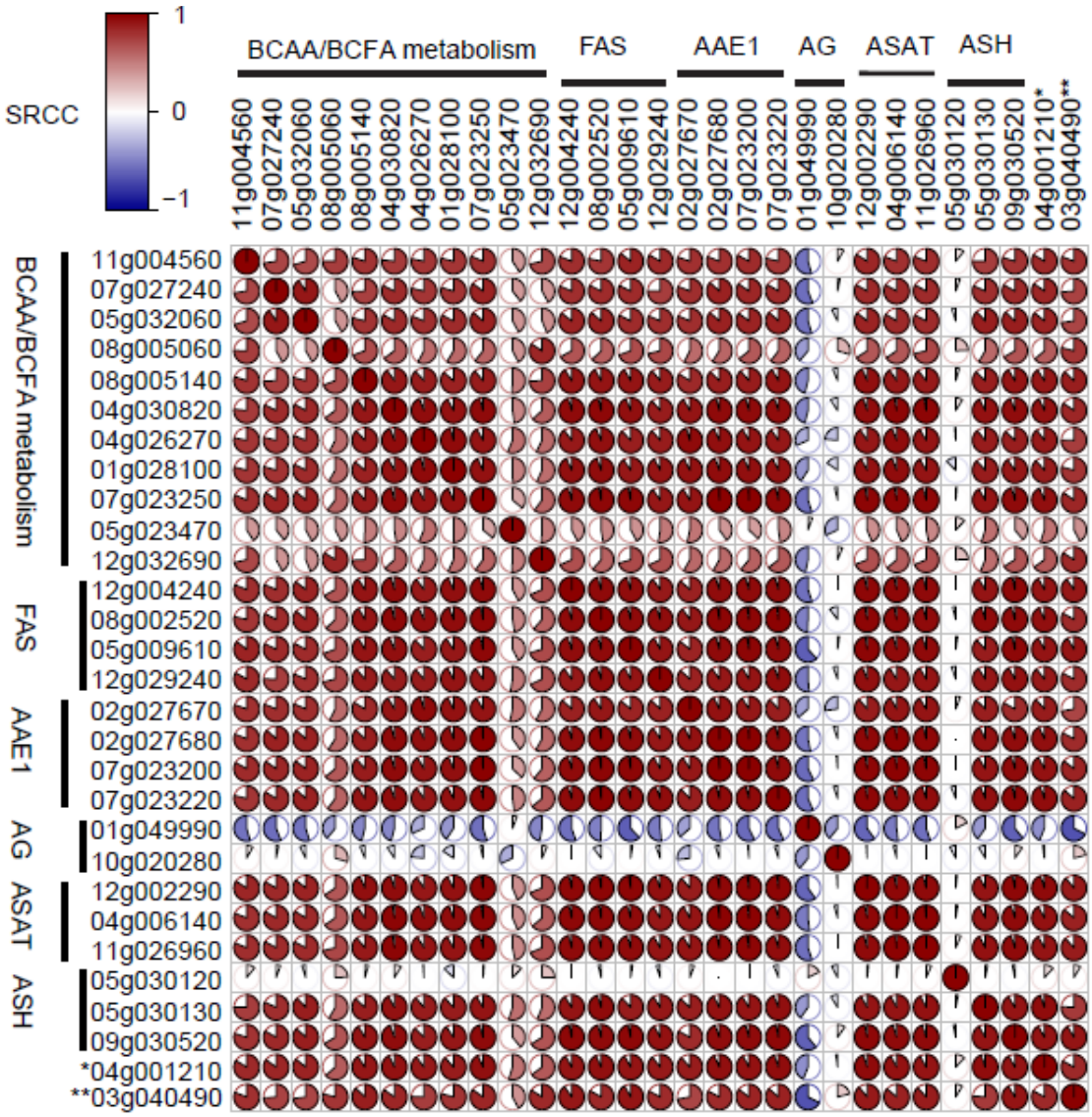

Figure 11. Correlation among expression profiles of selected genes. FPKM values for 29 phase 1- and phase 2-related genes in our 29 samples were used to determine pairwise Spearman's rank correlation coefficients (SRCC). A [29x29] matrix was created to visualize correlation. Red, white, and blue colors indicate positive, zero, and negative correlations, respectively. Darker colors indicate stronger correlations. *Sopen04g001210* (marked with an asterisk) is a carboxylesterase gene related to ASH. *Sopen03g040490* (marked with double-asterisk) is the recently identified invertase gene. Abbreviations: BCAA= branched-chain amino acid; FAS= fatty acid synthase components; AAE1= acyl-activating enzyme 1; AG= acylglucose phase 2 (previous model); ASAT= acylsucrose acyltransferase; ASH= acylsugar hydrolase.

2.3.5. Acylsugar Phase 2 Metabolic Genes are Upregulated in High-Acylsugar-Producing Accessions and Show Strong Positive Correlation with Phase 1 DEGs

Genes encoding acylsugar acyltransferases (ASATs) and the recently identified invertase (SpASFF1), which are involved in phase 2 of acylsugar biosynthesis (Schilmiller et al., 2015; Fan et al., 2016; Fan et al., 2017; Leong et al., 2019), were upregulated in high-acylsugar-producing accessions (Table 6). ASATs use acyl-CoA molecules as their substrates (Schilmiller et al., 2012; Schilmiller et al., 2015; Fan et al., 2016; Fan et al., 2017), and free straight-chain fatty acids (SCFAs) produced by FAS-catalyzed *de novo* fatty acid biosynthesis must be activated to their acyl-CoA derivatives (SCFA-CoA) before they can be used by ASATs (Figure 3A). All three ASAT genes exhibited strong positive correlation with both AAE1 genes and FAS components (ρ= 0.89 to 0.98; P = 1.23E-10 to <1E-16; Figure 11).

Recently, three genes encoding acylsugar hydrolases (ASHs; carboxylesterases that remove acyl groups from acylsucrose molecules) and a related carboxylesterase gene (*Sopen04g001210*) were reported in cultivated and wild tomato (Schilmiller et al., 2016). These genes, except *Sopen05g030120* (Sp-*ASH1*), were in the list of 1087 common DEGs (Table 4), and showed strong positive correlation with genes putatively involved in phase 1 acyl chain synthesis (Figure 11). The expression profile of *ASH1* in different tissues of *S. lycopersicum* led the authors to propose additional non-trichome-localized functions for Sl-ASH1 (Schilmiller et al., 2016); our results further support the idea that ASH1 may not have a critical role in trichome acylsucrose metabolism.

Table 6. Genes involved in phase 2 of acylsugar production.

Gene	'LOW' vs. 'HIGH'		LA1920 vs. LA0716		Annotation
	Log_2FC	**FDR**	**Log_2FC**	**FDR**	
Acylglucose biosynthesis based on biochemical reports (previous model of 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)
Recently identified invertase (current model of acylglucose biosynthesis)					
Sopen03g040490	2.99	9.60E-13	2.87	1.10E-11	Acylsucrose fructofuranosidase 1

Log_2FC and FDR indicate log_2 (fold-change) and false discovery rate (*P* values adjusted for multiple-testing), respectively. Positive and negative log_2FC values indicate higher and lower expression levels, respectively, in high-acylsugar-producing accessions.

2.3.6. Expression Levels of Two Previously Reported Acylglucose Biosynthetic Genes Are Not Consistent with Acylsugar Levels

Previous biochemical reports indicated that the first two steps in phase 2 of acylglucose biosynthesis are catalyzed by UDP-Glc:FA GT and SCPL GAT, respectively (Figure 3A) (Ghangas and Steffens, 1993; Kuai et al., 1997; Li et al., 1999;

Li and Steffens, 2000). *Sopen01g049990,* encoding the UDP-Glc:FA GT, showed 2-fold higher expression in the 'LOW' group (FDR= 2.70E-03), whereas *Sopen10g020280,* which encodes the SCPL GAT, had no significant difference in expression between low- and high-acylsugar-producing accessions (Table 6). Similarly, *Sopen01g049990* showed weak negative correlation with genes putatively involved in acyl chain synthesis (phase 1), whereas *Sopen10g020280* showed no significant correlation (Figure 11).

2.3.7. Three Genes Putatively Encoding ATP-Binding Cassette (ABC) Transporters Are Upregulated in High-Acylsugar-Producing Accessions

One important question in acylsugar metabolism is how acylsugar molecules are exported out of trichome cells. It remains unclear whether acylsugars are transported across the plasma membrane by membrane-associated transporter proteins or via vesicular transport. Five genes belonging to the ATP-binding cassette (ABC) transporter family showed differential expression, and three of them were upregulated in high-acylsugar-producing accessions (Table 7). Of these, two DEGs (*Sopen12g034820* and *Sopen04g005380*) belong to the G subfamily of ABC transporters. It has been suggested that ABC-G subfamily transporters, which are located predominantly on the plasma membrane and are involved in the deposition of surface lipids, may play important roles in herbivore defense by depositing insect-deterring specialized metabolites on leaf surfaces (Kang et al., 2011). Another gene (*Sopen03g001870*), predicted to encode an ABC-B subfamily transporter, had 23-fold higher expression in the 'HIGH' group (FDR= 1.10E-31).

Table 7. Differentially expressed genes involved in transport process, central carbon metabolism, and regulation of gene expression.

Gene	'LOW' vs. 'HIGH'		LA1920 vs. LA0716		Annotation
	Log_2FC	FDR	Log_2FC	FDR	
Transporters					
Sopen04g023150	-1.16	2.00E-02	-1.82	4.40E-09	ABC transporter F family
Sopen01g047950	-1.68	2.10E-02	-2.82	3.30E-04	ABC transporter G family
Sopen04g005380	3.19	2.80E-08	2.8	6.50E-04	ABC transporter G family
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
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- malic enzyme
Sopen12g026980	2.72	1.80E-09	1.21	3.40E-04	Glutathione S-transferase
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

Log_2FC and FDR indicate log_2 (fold-change) and false discovery rate (*P* values adjusted for multiple-testing), respectively. Positive and negative log_2FC values indicate higher and lower expression levels, respectively, in high-acylsugar-producing accessions. Transcription factor genes (marked with an asterisk) were selected based on their response to imazapyr treatment (see later).

2.3.8. Genes Involved in Central Carbon Metabolism Are Upregulated in High-Acylsugar-Producing Accessions

In high-acylsugar-producing accessions, acylsugar accumulation represents an astounding metabolic investment (up to 20% of leaf dry weight in LA0716) (Fobes et al., 1985; Shapiro et al., 1994). Because this extraordinarily high production of one class of specialized metabolites may require alteration of basic carbon metabolism, we investigated whether genes involved in central carbon metabolism were differentially expressed between low- and high-acylsugar-producing accessions. One gene (*Sopen07g006810*) putatively encoding the RUBISCO small subunit was upregulated in the 'HIGH' group (32-fold; FDR= 6.70E-12). Compared to other RUBISCO small subunit genes, which were not differentially expressed and had very high expression levels (FPKM as high as 5000), this differentially expressed gene had moderate levels of expression in medium- and high-acylsugar-producing accessions, and very low levels of expression in low-acylsugar-producing accessions (Figure 12).

Recently, the role of central carbon metabolism in supporting specialized metabolism in glandular trichomes of *S. lycopersicum* and *S. habrochaites* was investigated, and it was concluded that glandular trichomes import most of their fixed carbon from underlying leaf tissues, despite being able to perform some photosynthetic activities (Balcke et al., 2017). Special roles of chloroplast malic enzyme in supplying pyruvate and glutathione S-transferase in alleviating oxidative stress were also reported. We did find genes putatively encoding these central carbon metabolic proteins in the list of 1087 DEGs (Table 7; Figure 12).

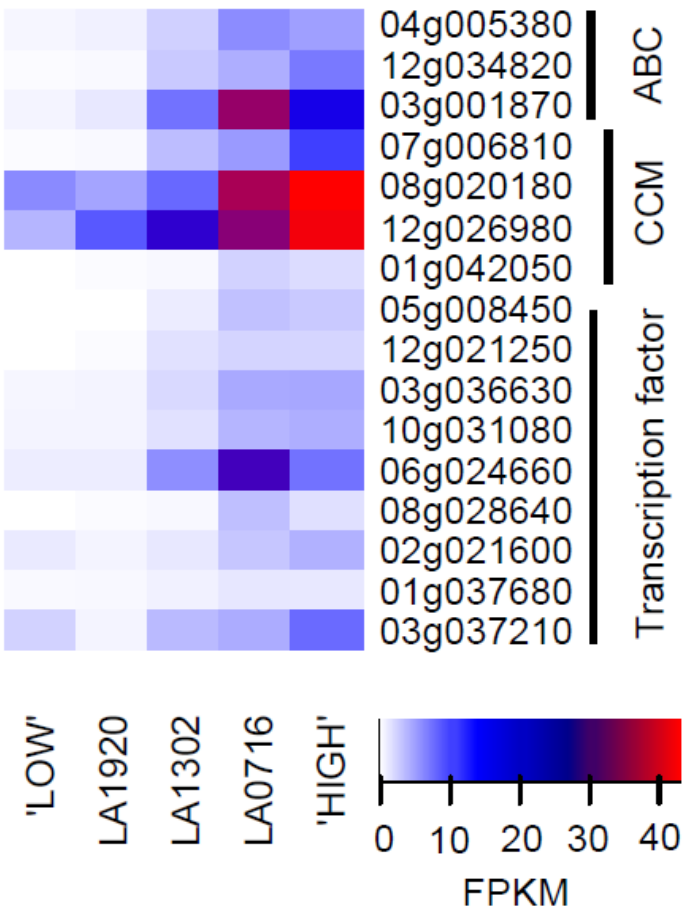

Figure 12. Heatmaps showing expression levels of genes with putative functions in acylsugar metabolism (Table 7 DEGs). Genes are designated by their gene identifier numbers (Sopen IDs). Two genes (*Sopen04g023150* and *Sopen01g047950*) were removed from the heatmap due to their very high expression levels. *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; ABC= ABC transporters; CCM= central carbon metabolism.

2.3.9. Trichome-Enriched Expression of Acylsugar Candidate Genes

Genes encoding ASATs and the recently identified invertase (SpASFF1) are expressed in apical cells of type I/IV glandular trichomes (Schilmiller et al., 2012; Schilmiller et al., 2015; Fan et al., 2016; Leong et al., 2019). Using data of Ning et al. (2015), who published expression profiles of *S. lycopersicum* genes in isolated stem trichomes versus shaved stems, we investigated if *S. lycopersicum* homologs of candidate genes show trichome-enriched expression pattern. We used reciprocal BLAST to identify putative *S. lycopersicum* orthologs of candidate genes, and they showed trichome-enriched expression pattern (26- to 562-fold for AAE1 members, 143- to 470-fold for ABC transporter genes, 16- to 287-fold for central carbon metabolic genes; Table 8).

Table 8. Trichome-enriched expression of acylsugar metabolic genes' putative orthologs in *S. lycopersicum*.

***S. pennellii* ID**	**Annotation**	***S. lycopersicum* ID**	**FC**
Fatty acid synthase (FAS) components and acyl-activating enzyme1 (AAE1) members			
Sopen12g004240	KAS IV/ KAS II-like domain	*Solyc12g009260*	102.7
Sopen08g002520	Beta-ketoacyl-ACP synthase III	*Solyc08g006560*	89.6
Sopen05g009610	Beta-ketoacyl-ACP reductase	*Solyc05g014560*	24.3
Sopen12g029240	Enoyl-ACP reductase domain	*Solyc12g056610*	420.0
Sopen02g027670	Acyl-activating enzyme 1	*Solyc02g082870*	300.2
Sopen02g027680	Acyl-activating enzyme 1	*Solyc02g082880*	26.3
Sopen07g023200	Acyl-activating enzyme 1	*Solyc07g043630*	561.5
Sopen07g023220	Acyl-activating enzyme 1	*Solyc07g043660*	294.4
ABC transporters, central carbon metabolic proteins, and transcription factors			
Sopen04g023150	ABC transporter F family member	*Solyc04g051800*	1.5

Table 8 continued.

***S. pennellii* ID**	**Annotation**	***S. lycopersicum* ID**	**FC**
Sopen01g047950	ABC transporter G family member	*Solyc01g105450*	0.8
Sopen04g005380	ABC transporter G family member	*Solyc04g010200*	160.9
Sopen12g034820	Pleiotropic drug resistance protein 1-like (ABC transporter G family)	*Solyc12g100190*	143.3
Sopen03g001870	ABC transporter B family member	*Solyc03g005860*	470.0
Sopen07g006810	RUBISCO small subunit	*Solyc07g017950*	46.8
Sopen08g020180	NADP-dependent malic enzyme	*Solyc08g066360*	16.0
Sopen12g026980	Glutathione S-transferase	*Solyc05g006730*	287.0
Sopen01g042050	Sugar transporter ERD6-like	*Solyc01g098490*	22.5
Sopen05g008450	AP2 domain	*Solyc05g013540*	252.4
Sopen12g021250	AP2 domain	*Solyc12g042210*	205.8
Sopen03g036630	AP2 domain	*Solyc03g117720*	24.2
Sopen10g031080	Homeobox associated leucine zipper	*Solyc10g080540*	1.3
Sopen06g024660	Myb-like DNA-binding domain	*Solyc06g066180*	0.0
Sopen08g028640	AP2 domain	*Solyc08g080290*	393.1
Sopen02g021600	Myb-like DNA-binding domain	*Solyc02g076670*	70.3
Sopen01g037680	TCP family transcription factor		
Sopen03g037210	Helix-loop-helix DNA-binding domain	*Solyc03g118310*	130.5

FC indicates fold-change values. Numbers indicate higher expression levels in isolated stem trichomes compared to underlying shaved stem, according to data of Ning et al. (2015).

2.3.10. Genes Putatively Encoding Oxylipin Metabolic Proteins, Subtilisin-Like Proteases, and Antimicrobial Defense Proteins Are Upregulated in Low-Acylsugar-Producing Accessions

Because acylsugars have important roles in plant defense (Gentile et al., 1968; Juvik et al., 1982; Goffreda et al., 1989; Hawthorne et al., 1992; Rodriguez et al., 1993; Juvik et al., 1994; Liedl et al., 1995), and defense response genes were present in the list

of 1087 DEGs (enriched GO term GO:0006952; Figure 4), we investigated defense DEGs. Nine genes putatively encoding fatty acid desaturase and lipoxygenase enzymes were differentially expressed, and eight of them were upregulated in low-acylsugar-producing accessions (Table 8). Among them, one particular lipoxygenase gene (*Sopen01g002520*) showed high expression levels in low-acylsugar-producing accessions, and nearly no expression in high-acylsugar-producing accessions (Figure 13), leading to very high levels of differential expression (254- and 1640-fold changes in 'LOW'/'HIGH' and LA1920/LA0716 comparisons, respectively; FDR= 2.40E-15 and 2.30E-40, respectively). Lipoxygenases catalyze the dioxygenation of polyunsaturated fatty acids produced by fatty acid desaturases, and this enzymatic conversion is the first step in the biosynthesis of a vast and diverse family of specialized metabolites, collectively known as oxylipins, which contribute to innate immunity (La Camera et al., 2004). Oxylipin response pathways are regulated by TGA transcription factors, and two classes of enzyme (oxo-phytodienoate reductase and glutathione S-transferase) have been reported to take important roles in detoxifying them or metabolizing them into less reactive molecules (Mueller et al., 2008). Genes putatively encoding these oxylipin signaling and metabolic proteins were also upregulated in low-acylsugar-producing accessions (Table 8).

Lipoxygenase-produced oxylipins have important roles in plant defense against herbivores (Halitschke and Baldwin, 2003) and microbial pathogens (Prost et al., 2005). Therefore, low-acylsugar-producing accessions may generate oxylipins to compensate for diminished antiherbivory and antimicrobial activities of acylsugars.

Consistent with this hypothesis, genes putatively encoding subtilisin-like proteases, which play important role in the recognition of plant pathogens and activation of immune responses (Figueiredo et al., 2014), and other proteins involved in plant defense response to fungal and other biotic stimulus were also upregulated in low-acylsugar-producing accessions (Table 8).

Table 9. Differentially expressed genes putatively involved in oxylipin-related and amine oxidase-related defense systems.

Gene	'LOW'/'HIGH'		LA1920/LA0716		Annotation
	Log_2 FC	FDR	Log_2 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
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
Other plant defense-related					
Sopen08g028270	-7.03	2.90E-16	-4.15	8.60E-05	Subtilisin-like protease precursor

Table 9 continued.

Gene	'LOW'/'HIGH'		LA1920/LA0716		Annotation
	Log_2 FC	FDR	Log_2 FC	FDR	
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

Log_2FC and FDR indicate log_2 (fold-change) and false discovery rate (*P* values adjusted for multiple-testing), respectively. Positive and negative log_2FC values indicate higher and lower expression levels, respectively, in high-acylsugar-producing accessions.

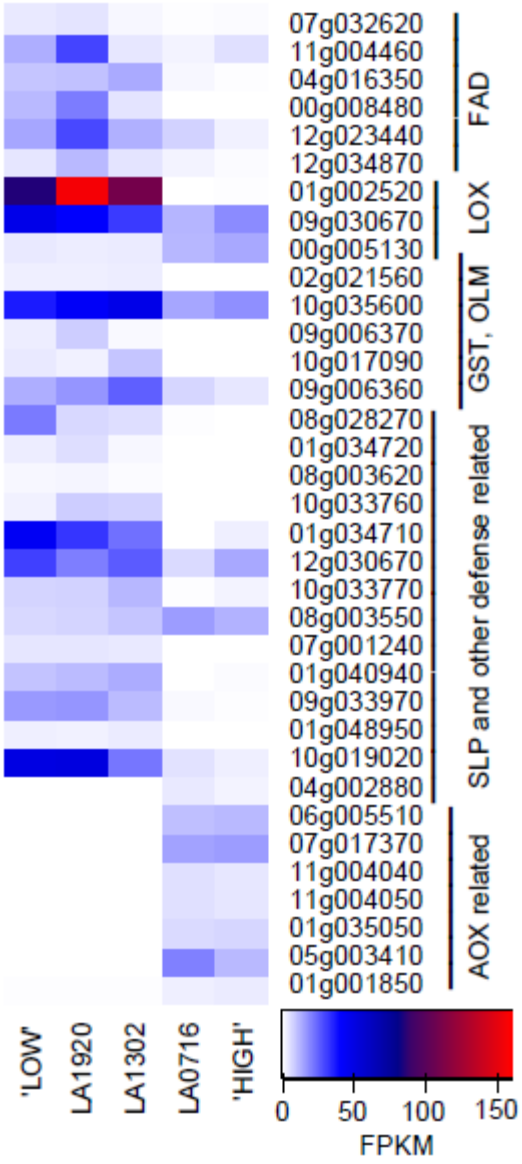

Figure 13. Heatmap showing expression levels of genes putatively involved in plant defense (Table 9 DEGs). Expression levels of many genes in LA1302 were similar to low-acylsugar-producing accessions but different from high-acylsugar-producing accessions. Some oxylipin-related defense genes, which were 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.

2.3.11. Amine Oxidase-Related Defense Response Genes Are Upregulated in High-Acylsugar-Producing Accessions

We found that one DEG (*Sopen06g005510*) predicted to encode the catalytic domain of glutamine synthetase had 183-fold higher expression in the 'HIGH' group than in the 'LOW' group (FDR= 2.60E-42). In plants, expression of glutamine synthetase is induced by higher concentrations of ammonia, which encouraged us to search for DEGs involved in ammonia-liberating reactions. Three genes encoding amine oxidase domains (*Sopen07g017370*, *Sopen11g004040*, and *Sopen11g004050*) showed 186-, 158-, and 66-fold higher expression, respectively in the 'HIGH' group (FDR= 3.20E-47, 9.20E-15, and 2.30E-14, respectively; Table 9). BLAST results indicated that *Sopen11g004040* and *Sopen11g004050* showed similarity with FLOWERING LOCUS D-like proteins, which have a flavin-containing amine oxidase domain, and are required for systemic acquired resistance in *Arabidopsis thaliana* (Singh et al., 2013). The three products of amine oxidase-catalyzed oxidative de-amination reactions are ammonia, H_2O_2, and an aldehyde that can be further oxidized to the corresponding carboxylic acid (Cona et al., 2006). One putative aldehyde oxidase gene (*Sopen01g035050*) was upregulated 229-fold in the 'HIGH' group (FDR= 5.00E-57).

H_2O_2 generated by amine oxidase- and aldehyde oxidase-catalyzed reactions acts as a signal molecule and plays an important role during pathogen invasion by acting on extensin proteins (Niebel et al., 1993; Jackson et al., 2001; Cona et al., 2006). Two genes predicted to encode extensin-like glycoproteins were upregulated in high-acylsugar-producing accessions (Table 9). Although amine oxidase-related defense genes were

upregulated in high-acylsugar-producing accessions, it is important to note that their expression levels were much lower than some oxylipin-related defense genes in low-acylsugar-producing accessions (Figure 13).

Using data of Ning et al. (2015), we checked if defense DEGs exhibit trichome-enriched expression. Putative *S. lycopersicum* orthologs of oxilipin-related defense DEGs do not show considerable trichome-enriched expression (except two out of 28 DEGs; *Sopen09g006360* and *Sopen10g019020*; Table 10). Among amine oxidase-related defense DEGs, we could identify putative *S. lycopersicum* ortholog (*Solyc01g005850*) for only one gene (*Sopen01g001850*; extension-like protein), and *Solyc01g005850* does not show trichome-enriched expression. Based on expression profiles in our 29 samples, genes involved in fatty acid desaturation, lipoxygenation, oxylipin signaling and metabolism, and antimicrobial defense showed moderate to strong correlation with each other (Figure 14). Similarly, amine oxidase-related defense DEGs also showed strong positive correlations among themselves. These results suggest that two different active networks of innate immunity were differentially expressed, although one network had considerably higher expression levels in low-acylsugar-producing accessions. Additionally, in contrast to acylsugar DEGs, defense DEGs do not exhibit trichome-enriched expression.

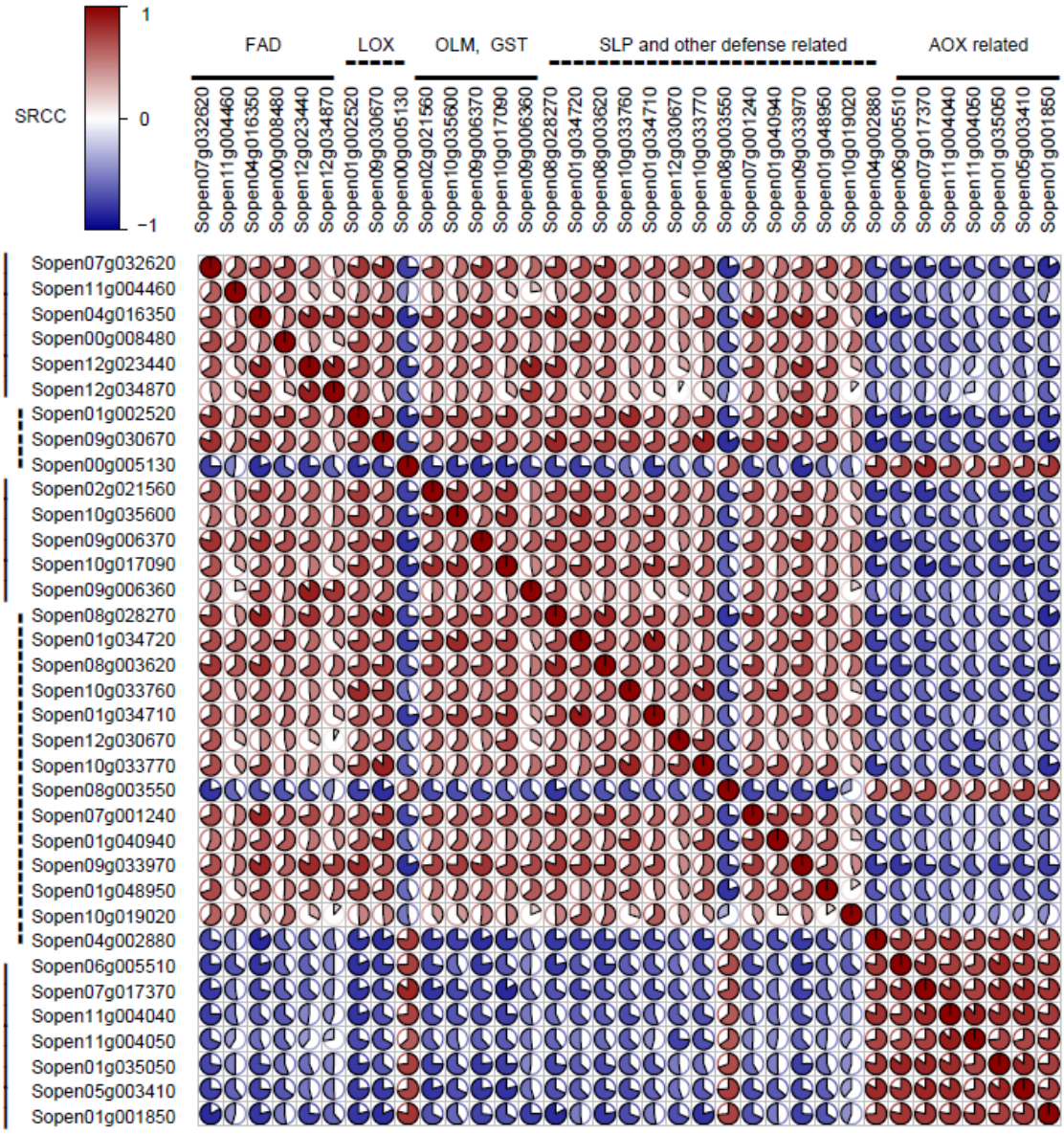

Figure 14. Correlation among expression profiles of putative defense DEGs. Expression profiles (FPKM values) for each of 35 genes (from Table 9) in our 29 samples were used to determine pairwise Spearman's rank correlation coefficient (SRCC). Red, white, and blue colors indicate positive, zero, and negative correlations, respectively. Darker colors indicate stronger correlations. Abbreviations: FAD= fatty acid desaturase; LOX= lipoxygenase; OLM= oxylipin metabolism; GST= glutathione S-transferase; SLP= subtilisin-like protease; AOX= amine oxidase.

Table 10. Trichome-enriched expression of defense genes' putative orthologs in *S. lycopersicum*.

***S. pennellii* ID**	**Annotation**	***S. lycopersicum* ID**	**FC**
Oxylipin-related defense genes			
Sopen07g032620	Fatty acid desaturase 4	*Solyc07g064410*	12.6
Sopen11g004460	Stearoyl-(ACP)-9-desaturase	*Solyc11g008680*	1.0
Sopen04g016350	Delta(12)-fatty acid desaturase	*Solyc04g040130*	0.1
Sopen00g008480	Delta(12)-fatty acid desaturase		
Sopen12g023440	Delta(12)-fatty acid desaturase	*Solyc12g049030*	0.1
Sopen12g034870	Delta(12)-fatty acid desaturase		
Sopen01g002520	Linoleate 13S-lipoxygenase		
Sopen09g030670	Linoleate 9S-lipoxygenase	*Solyc09g075860*	1.3
Sopen00g005130	Linoleate 9S-lipoxygenase		
Sopen02g021560	Transcription factor TGA5-like	*Solyc02g073570*	0.2
Sopen10g035600	12-Oxo-phytodienoate reductase	*Solyc10g086220*	1.4
Sopen09g006370	Glutathione S-transferase-like	*Solyc09g011600*	16.3
Sopen10g017090	Glutathione S-transferase		
Sopen09g006360	Glutathione S-transferase	*Solyc09g011590*	132.5
Sopen08g028270	Subtilisin-like protease precursor		
Sopen01g034720	Subtilisin-like serine protease SBT4	*Solyc01g087790*	9.1
Sopen08g003620	Subtilisin-like protease SBT1.7	*Solyc08g007670*	0.0
Sopen10g033760	Subtilisin-like protease SBT1.7	*Solyc10g084320*	0.0
Sopen01g034710	Subtilisin-like protease SBT1.7	*Solyc01g087770*	0.2
Sopen12g030670	Subtilisin-like protease SBT1.7	*Solyc12g088760*	0.0
Sopen10g033770	Subtilisin-like protease SBT1.7		
Sopen08g003550	Subtilisin-like protease SBT1.7		
Sopen07g001240	Chitotriosidase-1 (chitinase activity)	*Solyc07g005100*	4.1
Sopen01g040940	Wound-induced protein WIN1-like (chitin binding)	*Solyc01g097270*	0.9
Sopen09g033970	Pathogenesis-related protein STH-2	*Solyc09g091000*	0.1
Sopen01g048950	Pathogenesis-related protein 1A-like	*Solyc01g106600*	1.2
Sopen10g019020	Kirola-like defense response protein	*Solyc10g048030*	319.3
Sopen04g002880	Kirola-like defense response protein	*Solyc04g007770*	1.0

Table 10 continued.

***S. pennellii* ID**	**Annotation**	***S. lycopersicum* ID**	**FC**
Amine oxidase-related defense genes			
Sopen06g005510	Glutamine synthetase catalytic domain		
Sopen07g017370	Copper-containing amine oxidase		
Sopen11g004040	Protein FLOWERING LOCUS D (flavin-containing amine oxidase)		
Sopen11g004050	Protein FLOWERING LOCUS D (flavin-containing amine oxidase)		
Sopen01g035050	Aldehyde oxidase		
Sopen05g003410	Extensin-like protein		
Sopen01g001850	Extensin-like protein	*Solyc01g005850*	1.3

FC indicates fold-change values. Numbers indicate higher expression levels in isolated stem trichomes compared to underlying shaved stem, according to data of Ning et al. (2015).

2.3.12. Inhibition of BCAA Biosynthesis Changes Transcript Levels of Many Acylsugar Candidate DEGs, but Not Defense DEGs

Acetolactate synthase (ALS) catalyzes the first common step in the biosynthesis of BCAAs (Figure 7A), and inhibition of this enzyme significantly lowers acylsugar production in *S. pennellii* (Walters and Steffens, 1990). To determine whether inhibition of this enzyme changed expression of ALS and other candidate genes involved in acylsugar metabolism, we treated leaves of the high-acylsugar-producing accession LA0716 with the ALS inhibitor imazapyr at 0.1 mM and 1 mM concentrations. Many of the DEGs involved in BCAA/BCFA metabolism showed differential expression in response to imazapyr treatment (Table 11).

Table 11. Response of acylsugar metabolic genes to imazapyr treatment.

Gene	Imazapyr 0.1 mM		Imazapyr 1 mM		Annotation
	Log_2 FC	**FDR**	**Log_2 FC**	**FDR**	
Genes related to phase 1 of acylsugar biosynthesis (Table 4 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-Hdroxyisobutyryl-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 6 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

Table 11 continued.

Gene	Imazapyr 0.1 mM		Imazapyr 1 mM		Annotation
	Log_2 FC	**FDR**	**Log_2 FC**	**FDR**	
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
Sopen03g040490	-1.32	3.35E-04	-3.67	1.96E-23	Invertase (beta-fructofuranosidase)
Other genes related to acylsugar metabolism (Table 7 genes)					
Sopen04g023150	-0.01	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
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
Sopen12g026980	-1.12	2.36E-02	-4.14	8.94E-18	Glutathione S-transferase
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

Log_2FC and FDR indicate log_2 (fold-change) and false discovery rate (*P* values adjusted for multiple-testing), respectively. Positive and negative log_2FC values indicate higher and lower expression levels, respectively, in response to the inhibitor treatment compared to control solution.

Surprisingly, DEGs putatively involved in SCFA metabolism (FAS components and AAE1 members), showed significant downregulation at both 0.1 mM and 1 mM imazypyr in a concentration-dependent manner. It is interesting to note that AAE1 proteins have a peroxisomal location in *Arabidopsis thaliana* (Reumann et al., 2009); one gene (*Sopen11g007710*) predicted to encode Mpv17/PMP22 family peroxisomal membrane protein had 7.7-fold (FDR= 1.17E-08) and 165-fold (FDR= 3.62E-24) lower expression at 0.1 mM and 1 mM imazapyr treatment, respectively. *Sopen11g007710* was upregulated 29-fold in the 'HIGH' group (FDR= 2.80E-11), and its putative ortholog in *S. lycopersicum* (*Solyc11g012990*) has 1420-fold higher expression in isolated trichomes than shaved stems, consistent with trichome-enriched expression of AAE1 members (Ning et al., 2015). PMP22 is presumably involved in controlling permeability of peroxisomal membrane (Brosius et al., 2002). These results suggest a role of peroxisome in acylsugar metabolism.

DEGs encoding ASATs and the invertase (phase 2 of acylsugar biosynthesis) showed significant decrease in gene expression level in a concentration-dependent manner (by as much as 94-fold, FDR= 1.65E-14), as did ASHs and the related carboxylesterase *Sopen04g001210* (by as much as 70-fold, FDR= 8.82E-50) (Table 5). On the other hand, *Sopen01g049990* and *Sopen10g020280* (encoding UDP-Glc:FA GT and SCPL GAT, respectively; previous model of acylglucose biosynthesis) showed slightly higher and similar expression levels, respectively in response to imazapyr treatment.

DEGs putatively encoding central carbon metabolic proteins and three ABC transporters were significantly downregulated (by as much as 45-fold, FDR= 4.99E-10 for RUBISCO small subunit; Table 11). Imazapyr treatment in *Arabidopsis thaliana* resulted in significant induction of nine genes (out of the ten genes reported) encoding ABC transporters, indicating their role in detoxification process (Manabe et al., 2007). Similarly, 15 putative ABC transporter genes in *S. pennellii* were upregulated in response to imazapyr either at higher concentration or at both concentrations (Table 12). However, three putative ABC transporter DEGs were repressed by imazapyr treatment, consistent with expression profiles of other acylsugar metabolic genes.

Table 12. Effect of imazapyr treatment on expression levels of ABC transporter genes.

	Imazapyr 0.1 mM		**Imazapyr 1 mM**		
Gene	**Log_2FC**	**FDR**	**Log_2FC**	**FDR**	**Annotation**
Sopen02g020400	1.078	0.006991	4.217	4.38E-33	ABC transporter B family
Sopen03g003490	1.151	0.00018	1.901	7.16E-12	ABC transporter C family
Sopen06g025860	1.024	0.002743	1.656	3.81E-08	ABC transporter I family
Sopen08g031080	0.421	0.226405	1.459	7.11E-08	ABC transporter F family
Sopen04g024980	0.245	0.506156	1.346	1.79E-07	ABC transporter D family
Sopen01g044530	1.200	0.185844	3.320	1.88E-07	Pleiotropic drug resistance protein
Sopen06g017680	-0.301	0.383764	1.259	4.36E-07	Putative ABC1 protein At2g40090
Sopen09g017780	0.470	0.126481	1.226	6.28E-07	ABC transporter B family
Sopen01g043640	0.459	0.137201	1.189	1.33E-06	ABC transporter F family
Sopen06g012460	-0.476	0.187344	1.182	1.69E-05	ABC transporter C family
Sopen06g031340	-0.063	0.89446	1.079	5.21E-05	ABC transporter F family

Table 12 continued.

	Imazapyr 0.1 mM		Imazapyr 1 mM		
Gene	**Log_2FC**	**FDR**	**Log_2FC**	**FDR**	**Annotation / domain**
Sopen11g027810	0.279	0.520061	1.222	6.96E-05	ABC transporter F family
Sopen08g023990	0.260	0.656472	1.267	0.001124	ABC transporter E family
Sopen11g026980	0.007	1	1.900	0.002469	ABC transporter B family
Sopen01g047950	0.390	0.551545	1.265	0.006381	ABC transporter G family
DEGs which had higher exression levels in low-acylsugar-producing accessions					
Sopen04g023150	-0.014	0.987791	-0.243	0.511372	ABC transporter F family
Sopen01g047950	0.390	0.551545	1.265	0.006381	ABC transporter G family

Unlike known and candidate acylsugar metabolic DEGs, the amine oxidase-related defense DEGs did not show any concentration-dependent downregulation in response to ALS inhibition (Table 13). These results suggested that additional genes involved in acylsugar metabolism may also be present in the list of 171 DEGs (out of 1087 common DEGs between low- and high-acylsugar-producing accessions) that responded to imazapyr at both concentrations. Of the 13 transcription factor DEGs, nine were present in this list (Table 7), and putative *S. lycopersicum* orthologs of six of them show trichome-enriched expression profile (Table 8) (Ning et al., 2015).

Table 13. Effect of imazapyr treatment on expression levels of amine oxidase-related defense genes.

Gene ID	Imazapyr 0.1 mM		Imazapyr 1 mM		Annotation
	Log_2FC	FDR	Log_2FC	FDR	
Sopen06g005510	-0.82	0.10	-1.65	8.00E-05	Glutamine synthetase domain
Sopen07g017370	0.31	0.41	1.20	1.01E-05	Copper-containing amine oxidase
Sopen11g004040	-0.26	0.65	0.24	0.58	Protein FLOWERING LOCUS D (flavin-containing amine oxidase)
Sopen11g004050	-0.31	0.57	0.37	0.37	Protein FLOWERING LOCUS D (flavin-containing amine oxidase)
Sopen01g035050	-0.20	0.67	-0.04	0.92	Aldehyde oxidase
Sopen05g003410	0.54	0.07	0.22	0.43	Extensin-like protein
Sopen01g001850	-0.35	0.66	1.08	0.02	Extensin-like protein

2.3.13. Identification of Co-Expressed Gene Network Associated with Acylsugar Metabolism

Using expression profiles of *S. pennellii* genes in our 38 samples (29 from different accessions and 9 from the imazapyr treatments), weighted gene correlation network analysis (WGCNA) (Langfelder and Horvath, 2008) identified 41 co-expressed gene modules. Known and candidate acylsugar metabolic genes were clustered in the 'darkorange' module (182 genes; Figure 15A). One-third of the genes that responded to imazapyr (57/171) were also placed in this module. We selected the top-50 most strongly connected genes (based on intramodular connectivity as determined by WGCNA) for each of the three ASAT genes and the invertase gene, and merged them to identify the most strongly interconnected gene network. This network included genes involved in BCAA/BCFA metabolism, SCFA synthesis (FAS components) and their

activation (AAEs and the related peroxisomal membrane protein PMP22/Mpv17), acylsucrose metabolism (three ASATs, ASH2, ASH3 and the related carboxylesterase *Sopen04g001210*), as well as genes putatively encoding three ABC transporters, central carbon metabolic proteins, six transcription factors (including three AP2-family transcription factors) and other proteins (Figure 15B).

However, not all genes in this network had an obvious connection to acylsugar metabolism. One gene on chromosome 11 (*Sopen11g003320*; UDP-glucose:catechin glucosyltransferase (Noguchi et al., 2008)) and three sequential genes on chromosome 6 (*Sopen06g034810*, *Sopen06g034820*, and *Sopen06g034830*; myricetin methyltransferase (Schmidt et al., 2012; Kim et al., 2014)) showed strong intramodular connectivity with acylsugar metabolic genes. These flavonoid metabolic genes were also strongly downregulated in response to imazapyr treatment (Figure 16; Table 14), and their putative orthologs in *S. lycopersicum* are preferentially expressed in trichomes (Ning et al., 2015). In addition to abundant acylsugar compounds, flavonoid compounds, such as methylated myricetin, have been reported in *S. pennellii* trichomes (McDowell et al., 2011). Together, these results suggest that metabolisms of acylsugar and flavonoid compounds are strongly connected.

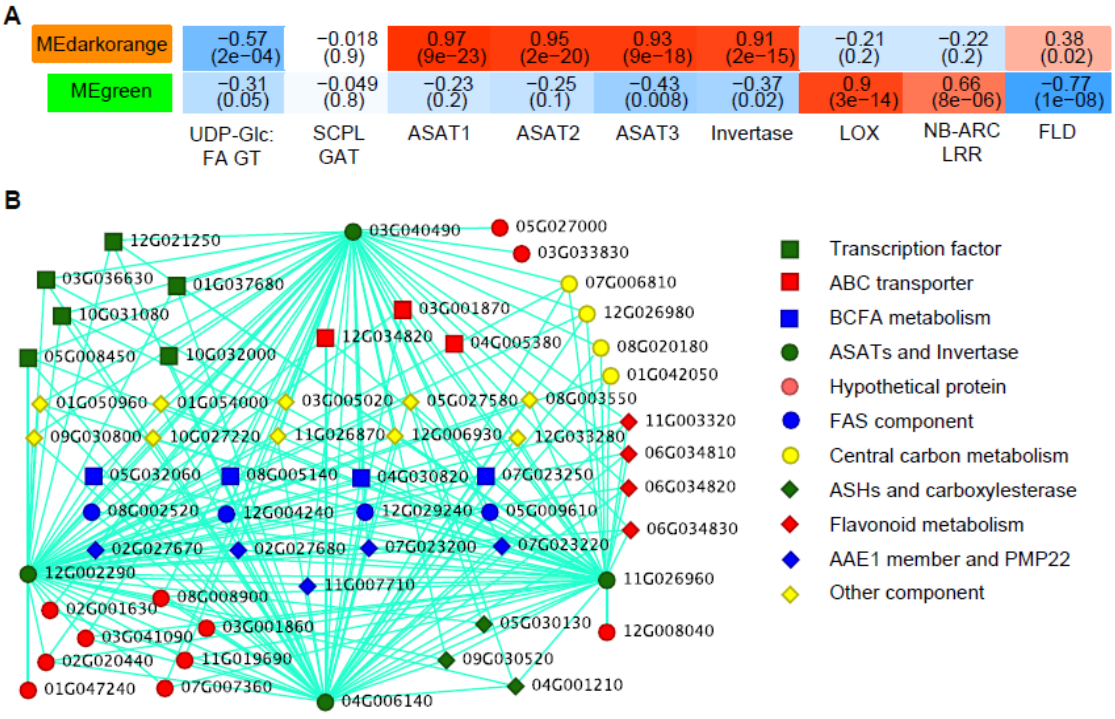

Figure 15. Weighted gene correlation network analysis (WGCNA) (Langfelder and Horvath, 2008). (A) Association of module eigengenes (MEs) with expression profiles of selected nine genes. WGCNA identified 41 modules; for each module, a module eigengene was selected as representative of the expression profile for that module (first priniciple component). Numbers indicate Spearman's rank correlation coefficient (SRCC) values between MEs of two selected modules and nine genes' expression profiles. *P* values associated with SRCC are given in parentheses. Blue, white, and red colors indicate negative, zero, and positive correlations, respectively. Acylsugar metabolic genes (ASATs and invertase) and lipoxygenase (LOX)-correlated genes were clustered in 'darkorange' and 'green' modules, respectively. NB-ARC-LRR indicates a recently characterized sequence (*Sopen09g036080*) that confers resistance against tospoviruses in *S. lycopersicum* and wild tomato species (Zhu et al., 2017). (B) 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 three ASAT genes and the invertase gene. These four top-50 lists contained only 58 different genes.

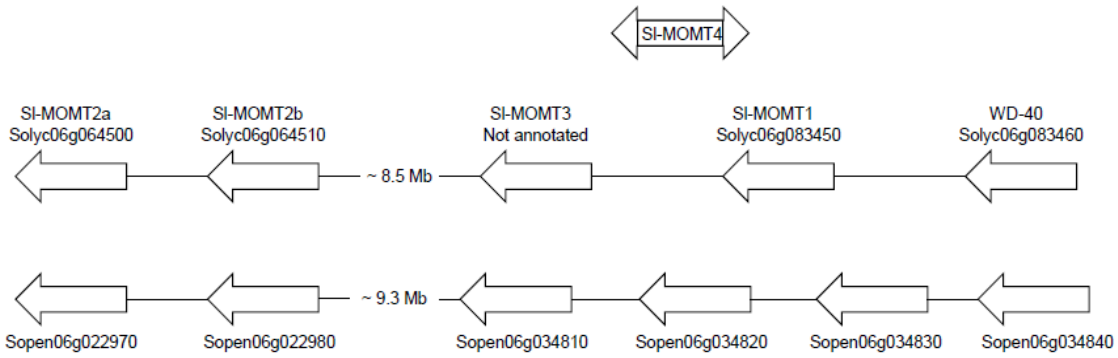

Figure 16. Selected *O*-methyltransferase genes on chromosome 6 involved in flavonoid metabolism in *S. pennellii* and *S. lycopersicum*. Homologous regions in *S. lycopersicum* and *S. pennellii* containing MOMT (myricetin *O*-methyltransferase) genes. Sl- *MOMT4* is not present in *S. lycopersicum* Heinz 1706 reference genome; in M82 cultivar, Sl-*MOMT4* is located near Sl- *MOMT1*, but not directly adjacent to Sl-*MOMT1* (Kim et al., 2014).

Table 14. Effect of imazapyr treatment on expression levels of flavonoid metabolic genes.

Gene	**Imazapyr 0.1 mM**		**Imazapyr 1 mM**		**Annotation**
	Log_2FC	**FDR**	**Log_2FC**	**FDR**	
Sopen06g022970	-0.57	4.82E-01	-2.7	6.63E-05	O-methyltransferase
Sopen06g022980	-1.61	9.31E-02	-2.8	6.87E-04	O-methyltransferase
Sopen06g034810	-1.61	1.27E-02	-5.32	1.59E-14	O-methyltransferase
Sopen06g034820	-1.72	5.07E-03	-6.29	3.80E-18	O-methyltransferase
Sopen06g034830	-2.58	4.54E-04	-8.38	1.36E-15	O-methyltransferase
Sopen06g034840	-0.25	4.73E-01	0.54	3.57E-02	WD domain repeat

Differential expression of *O*-methyltransferase genes in response to imazapyr treatment. Abbreviations: FDR= false discovery rate; logFC= log_2(fold-change).

2.3.14. Plant Defense Gene Network

We selected *Sopen01g002520* (lipoxygenase) as a representative of the oxylipin-mediated defense system, due to its robust differential expression (Figure 13; Table 9), and WGCNA identified co-expressed genes in the 'green' module. Enriched GO terms in this module include 'defense response' (GO:0006952), 'oxylipin biosynthetic process (GO:0031408), and 'glutathione metabolic process' (GO:0006749) (Figure 17). In the plant innate immune system, recognition and successful prevention of pathogen invasion rely on a series of molecular events, including pathogen detection at cell-surface by receptor kinases, a series of phosphorylation events with cytoplasmic kinases, nucleotide binding, release of intracellular calcium, protein ubiquitination, and transcriptional reprogramming (Couto and Zipfel, 2016). GO terms associated with these processes were significantly enriched in the 'green' module (FDR= 8.83E-17 to 4.6E-02). 20 of 104 expressed *S. pennellii* NB-ARC sequences and 23 of 150 leucine-rich repeat (LRR) sequences were clustered in the 'green' module, including recently characterized NB-ARC-LRR sequence Sw-5b (*Sopen09g036080*) that confers broad-spectrum resistance against American-type tospoviruses in *S. lycopersicum* and wild tomato species (Zhu et al., 2017).

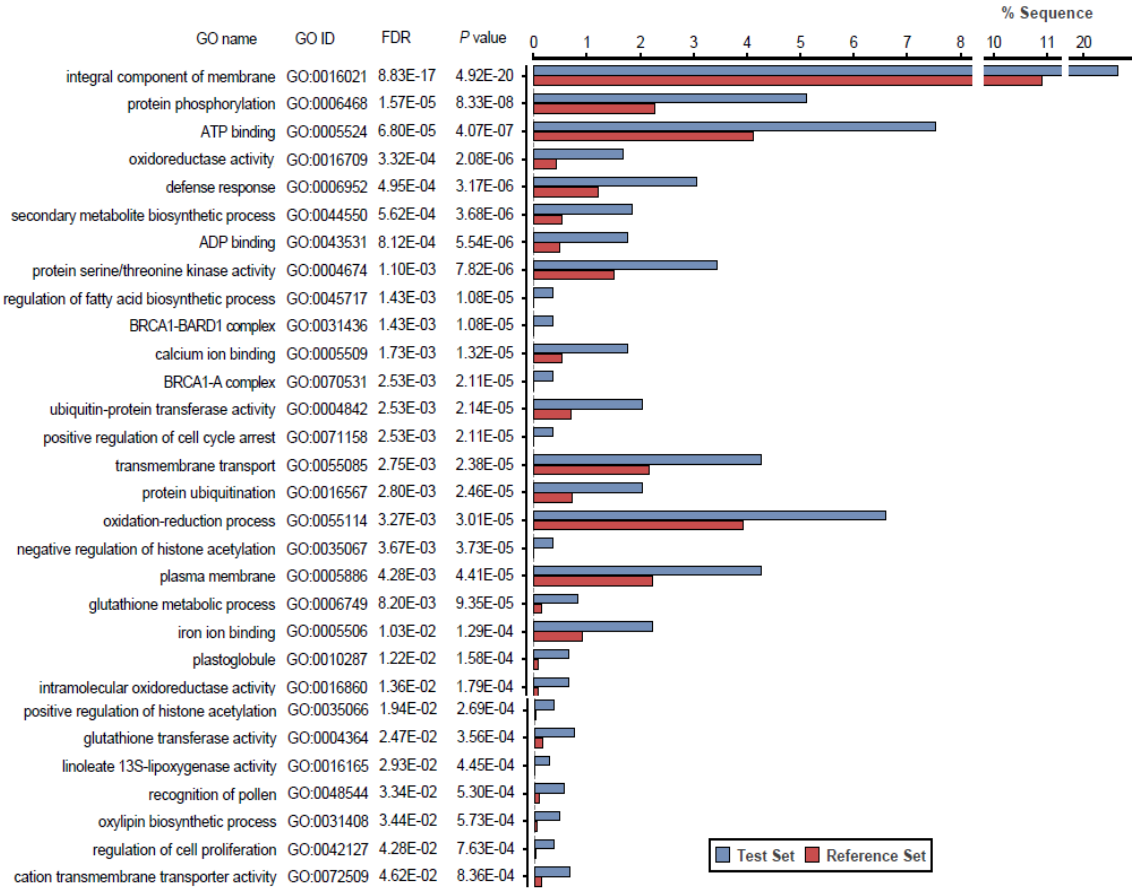

Figure 17. Enrichment of selected gene ontology (GO) terms associated with WGCNA 'green' module genes. Test set and reference set indicate 'green' module genes and *S. pennellii* annotated sequences, respectively. Enriched sequences include those putatively involved in oxylipin metabolism and plant defense signaling (Couto and Zipfel, 2016). Enrichment analysis was performed using Blast2GO software (Fischer's exact test). Only significant GO terms are shown (false discovery rate, FDR < 0.05).

2.3.15. Functional Validation of Two Candidate Genes Involved in SCFA Biosynthesis

To our knowledge, there are no reports of genes involved in acylsugar SCFA synthesis. We selected two candidate genes for functional validation (*Sopen08g002520* and *Sopen12g004240*; predicted to encode KAS III and KAS IV/KAS II-like enzymes,

respectively). These genes have 235- and 60-fold, respectively, higher expression levels in isolated stem trichomes compared to underlying tissue (Figure 18). In fact, *Sopen08g002520* showed a Sp*ASAT1*-like trichome-enriched expression profile. We identified *Solyc08g006560* and *Solyc12g009260* as putative *S. lycopersicum* orthologs of *Sopen08g002520* and *Sopen12g004240*, respectively, and these two tomato orthologs also show trichome-enriched expression profiles (90- and 103-fold, respectively; Table 8; Ning et al. (2015)). Phylogenetic analyses supported these orthologous relationships and revealed distinct clades of KAS III and KAS II sequences in solanaceous species (Figure 19 and 20).

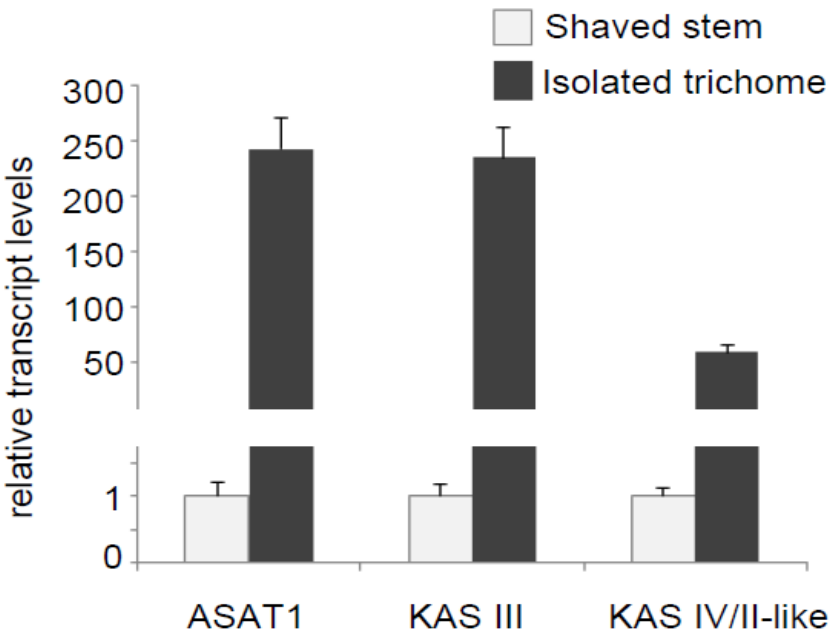

Figure 18. Trichome enriched expression of two candidate genes. Transcripts levels in isolated stem trichomes and underlying tissues (shaved stems) were measured by qRT-PCR. Error bars indicate SE (n = 5 biological replicates).

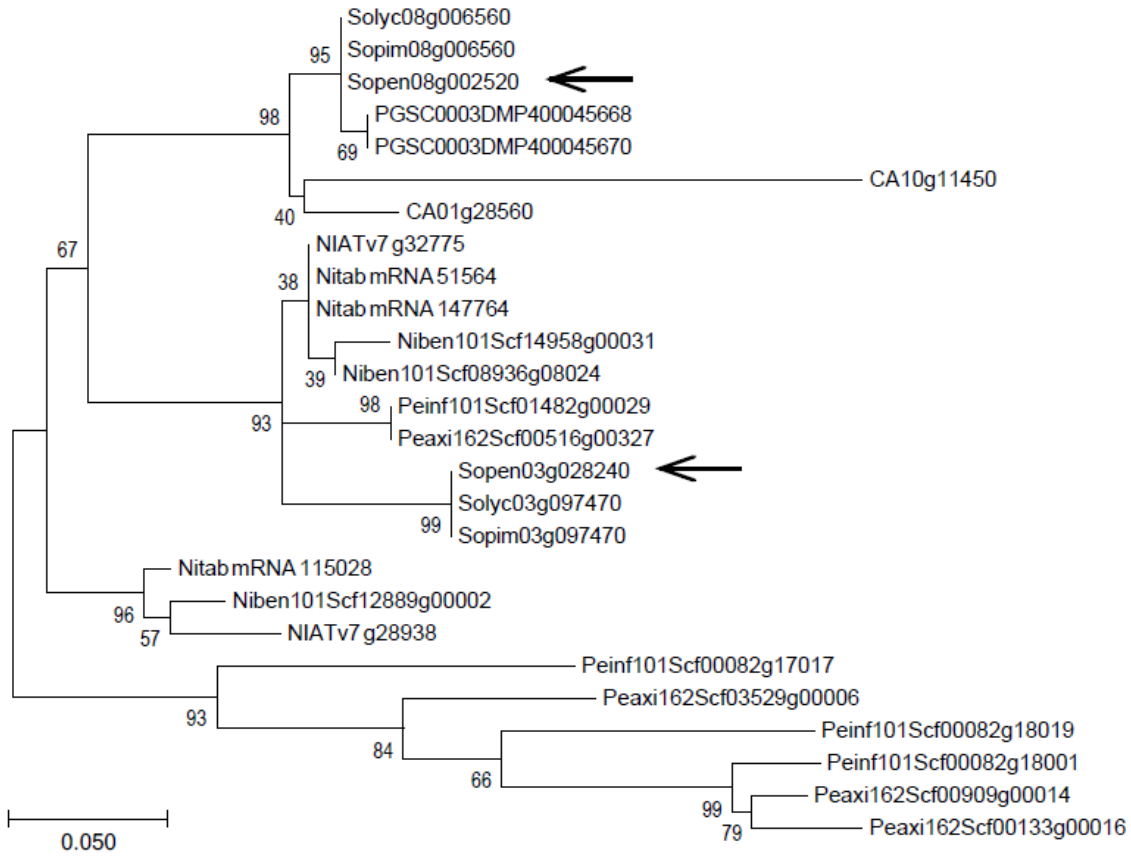

Figure 19. Phylogenetic relationships among KAS III sequences from different solanaceous species. Predicted transit peptide sequences were removed from protein sequences before multiple sequence alignment using MUSCLE (Edgar, 2004). The evolutionary history was inferred by using the maximum likelihood method based on the JTT+G model (Jones et al., 1992) (lowest Bayesian Information Criterion value). Bootstrap values from 1000 replicates are shown on the nodes. All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA7 (Kumar et al., 2016). KAS III sequences from *Solanum pennellii* are indicated with arrows. Sopen= *Solanum pennellii*, Sopim= *Solanum pimpinellifolium*, Solyc= *Solanum lycopersicum*, PGSC= Potato (*Solanum tuberosum*), CA=*Capsicum annuum*, Niben= *Nicotiana benthamiana*, Nitab= *Nicotiana tabacum*, NIAT= *Nicotiana attenuata*, Peaxi= *Petunia axillaris*, Peinf= *Petunia inflata*.

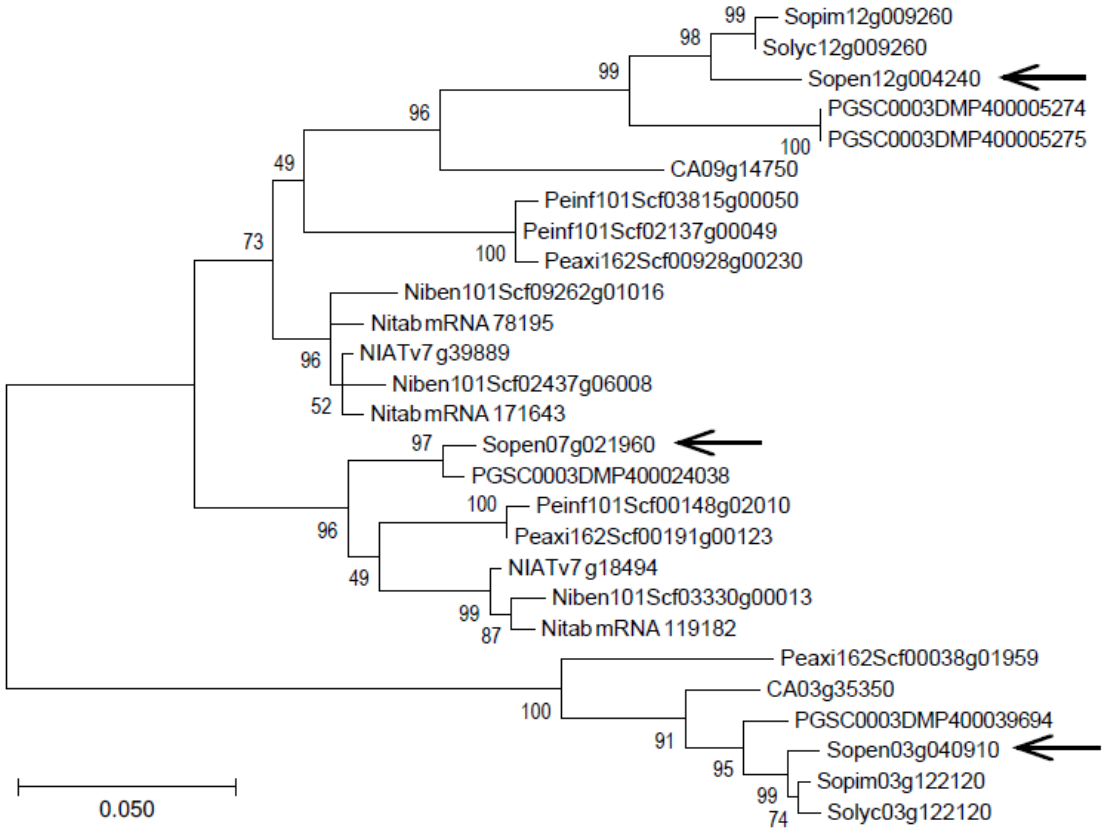

Figure 20. Phylogenetic relationships among KAS II sequences from different solanaceous species. KAS II sequences from *Solanum pennellii* are indicated with arrows.

We targeted these two *S. pennellii* loci in LA0716 for VIGS using tobacco rattle virus (TRV)-based silencing vectors, and acylsugar SCFA production was reduced by up to 40% in both *n*-C10 and *n*-C12 amount ($P < 0.001$; Figure 21A). VIGS resulted in significant downregulation of target genes (60% and 67% reduction in transcript levels for *Sopen08g002520* and *Sopen12g004240*, respectively; Figures 21B and C). Residual expression presumably is due to the fact that VIGS in *S. pennellii* is incomplete and inconsistent (Figure 21D). Additionally, no morphologically distinct areas would allow

us for better area-selective metabolite profiling, since no significant morphological differences were observed between silenced plants and control plants. Nevertheless, our results indicate that the two candidate genes, which are preferentially expressed in trichomes, play important roles in acylsugar SCFA biosynthesis. This is consistent with the previous prediction that KAS enzyme(s) other than KAS I might be involved in SCFA production in *S. pennellii* acylsugars (Slocombe et al., 2008).

2.4. Discussion

Acylsugars are powerful natural pesticides (Puterka et al., 2003), and increasing acylsugar-mediated insect pest resistance has been a target of various breeding programs in solanaceous crops (Bonierbale et al., 1994; Lawson et al., 1997; Alba et al., 2009; Leckie et al., 2012). The success of these breeding programs and bioengineering of acylsugar production will largely depend on unraveling the network of structural and regulatory genes. Here, we used transcriptomics approaches to identify candidate acylsugar metabolic gene network in *S. pennellii*. VIGS of two candidate genes provided functional validation of the gene network. Additionally, differential gene expression analysis and co-expression network analysis identified defense genes in *S. pennellii*.

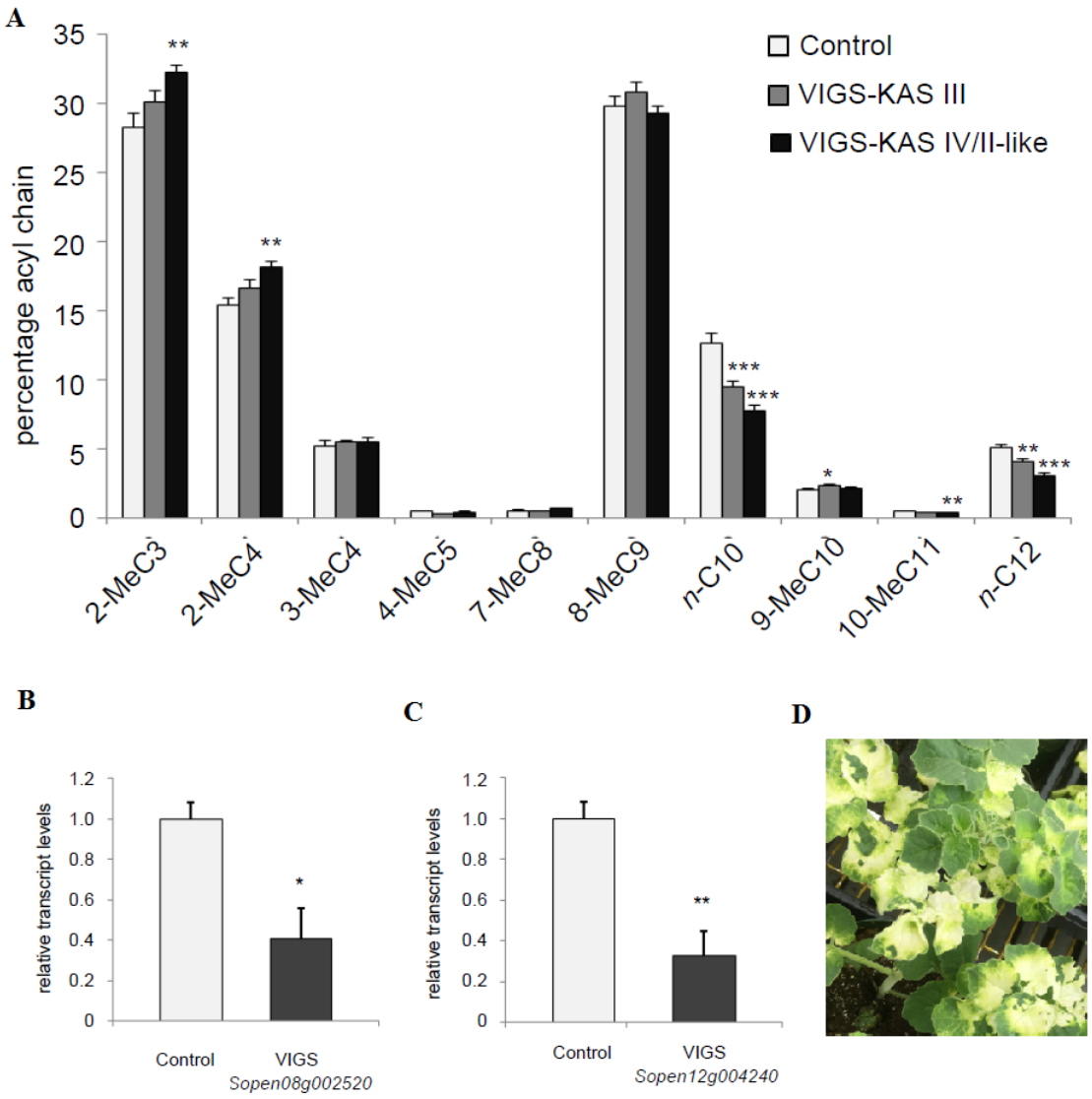

Figure 21. Virus-induced gene silencing of two candidate genes. (A) Acyl chain composition of acylsugars extracted from leaves of control plants (empty TRV vector) and two other groups of plants in which *Sopen08g002520* (KAS III) and *Sopen12g004240* (KAS IV/II-like), respectively, were targeted for VIGS. Relative abundance of acyl chains were measured by gas chromatography-mass spectrometry (GC-MS). Asterisks denote statistical significance compared to control plants as determined by one-way ANOVA (Dunnett's test; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). Error bars indicate SE (n = 10 biological replicates). (B) and (C) Downregulation of *Sopen08g002520* and *Sopen12g004240*, respectively. Transcripts levels were measured by qRT-PCR in control plants and in plants in which genes were targeted for VIGS. Error bars indicate SE (n = 3 biological replicates). (D) VIGS of a phytoene desaturase gene (*Sopen03g041530*), which was used as a positive control to assess the efficiency of VIGS.

2.4.1. Defense Pathway Genes

Enriched GO terms associated with pathogen recognition and downstream signaling cascades in the 'green' module indicate an active defense network in low-acylsugar-producing accessions (Figure 17). Oxylipins (Halitschke et al., 2003; Prost et al., 2005) and subtilisin-like proteases (Figueiredo et al., 2014) have important roles in plant defense, and higher expression levels of these defense genes predominantly in low-acylsugar-producing accessions (Figure 13) may indicate a compensatory mechanism for diminished defense activities of acylsugars in these accessions. Alternatively, expression of these defense genes may be suppressed in high-acylsugar-producing accessions due to the protection provided by acylsugars. Our data do not allow us to distinguish between these two possibilities, but it is important to note that all plants used in this study were grown in the same growth chamber, and no insects or symptoms of disease were apparent on any plant.

A different class of defense genes, including those putatively encoding amine oxidases, aldehyde oxidase, and extensin-like proteins, was upregulated in high-acylsugar-producing accessions. In general, these amine oxidase-related DEGs had much lower expression levels than some oxylipin-related defense DEGs in low-acylsugar-producing accessions (Figure 13). These different levels of expression may indicate that defense DEGs in high-acylsugar-producing accessions are involved in baseline innate immunity, whereas acylsugars provide the major defensive barrier. During pathogen or herbivore attack, these amine oxidases may be upregulated to produce H_2O_2, which can activate extensin proteins (Cona et al., 2006). Increased levels of extensin transcripts

upon nematode infection has been reported in tobacco (Niebel et al., 1993), and deposition of extensin mediated by H_2O_2-induced peroxidase during callus culture elicitation has been reported in *Vitis vinifera* (Jackson et al., 2001).

Most defense DEGs (both oxylipin-related and amine oxidase-related) had similar expression levels in the 'MEDIUM' accession LA1302 and in low-acylsugar-producing accessions (Figure 13). Multi-dimensional scaling plots also showed that the overall gene expression profile of LA1302 is similar to that of low-acylsugar-producing accessions, but very different from that of high-acylsugar-producing accessions (Figures 5A and 5B). These results are consistent with the prediction that LA1302 is more closely related to low-acylsugar-producing accessions (Shapiro et al., 1994). Despite the similarity of defense DEGs expression and the overall transcriptome profile between LA1302 and low-acylsugar-producing accessions, many DEGs with known and putative functions in acylsugar metabolism showed intermediate expression levels in LA1302 (Figures 10 and 12). Together, these results suggest that there is no strict linear relationship between acylsugar accumulation and expression of defense pathways.

2.4.2. Acylsugar Metabolic Genes

Our initial hypothesis for this work was that expression levels of previously reported acylsugar biosynthetic genes would be consistent with acylsugar levels in different accessions of *S. pennellii*, and that additional, unknown genes involved in acylsugar metabolism would be coordinately expressed with them. Our approach was validated by the identification of several DEGs with known functions in acylsugar

biosynthesis, such as BCAA/BCFA metabolic genes and acylsucrose biosynthetic genes (ASATs). Next, we used a chemical genetics approach to refine the list of acylsugar candidate genes by treating *S. pennellii* leaves with imazapyr to inhibit BCAA biosynthesis. This approach allowed us to separate the acylsugar metabolic DEGs from defense DEGs, which did not respond to imazapyr treatment. While we expected that genes related to synthesis of acylsugar BCFAs (derivatives of BCAAs) would be downregulated as a result of lack of substrate availability in response to imazapyr, our results showed that genes related to SCFA metabolism and other aspects of acylsugar metabolism (including known acylsucrose metabolic genes and candidate genes) also responded to imazapyr in a dose–dependent manner. It is important to note that the repression of acylsugar metabolic genes we observed after imazapyr treatment is distinct from the induction of general stress response genes that can result from amino acid starvation (Zhao et al., 1998).

Strong correlations among expression profiles of selected acylsugar metabolic genes (Figure 11) encouraged us to perform gene co-expression network analysis, which identified candidate gene networks involved in acylsugar metabolism and also plant defense (Figure 15 and 17). Additionally, data of Ning et al. (2015) indicate that putative *S. lycopersicum* orthologs of most of our acylsugar candidate genes, but not defense DEGs, are preferentially expressed in trichomes (Tables 8 and 10). Together, these results strongly support the candidacy of novel acylsugar metabolic genes that we have identified; candidate genes include those encoding KAS enzymes and other FAS components (for SCFA synthesis), AAE1 members (for activating SCFAs for ASATs), a

PMP22/Mpv17 family peroxisomal membrane protein (for controlling permeability of peroxisomal membrane), ABC transporters (for acylsugar transport), and central carbon metabolic proteins (for providing carbon to support acylsugar production), including a glutathione S-transferase (for alleviating oxidative stress).

In plant primary metabolism, the condensation reactions of fatty acid biosynthesis are catalyzed by KAS enzymes of the FAS complex, and at least three different KAS isoenzymes (KAS I, KAS II, and KAS III) are involved in *de novo* fatty acid biosynthesis. We selected two candidate *KAS* genes that are preferentially, if not exclusively, expressed in *S. pennellii* trichomes, and knockdown of these two novel genes led to significant decrease in SCFA synthesis (Figure 21), demonstrating their role in acylsugar metabolism. Duplication of primary metabolic gene followed by neofunctionalization (Ning et al., 2015) and spatial gene regulation have important functions in regulating trichome metabolism of acylsugar molecules, and KAS sequences also appear to follow that trend.

2.4.3. Regulation of Acylsugar Production

Both comparative transcriptomics and imazapyr treatment studies revealed that transcript levels for the two previously reported phase 2 acylglucose biosynthetic enzymes (UDP-Glc:FA GT and SCPL GAT) were not positively correlated with acylsugar levels (Tables 6and 11). Recently, Leong et al. (2019) clearly demonstrated role of the trichome-expressed invertase (SpASFF1) in acylglucose biosynthesis in *S. pennellii* trichomes. Their results contradict involvement of UDP-Glc:FA GT and SCPL

GAT in biosynthesis of acylglucose molecules, and our results are consistent with this finding. Our results also suggest that a majority of the acylsugar metabolic pathway, from acyl chain synthesis to acylsugar secretion, represents a remarkable co-expressed genetic network. WGCNA revealed that six transcription factor genes form a strong gene co-expression network with many acylsugar and flavonoid metabolic genes (Figure 15). Downregulation of these transcription factor genes in response to imazapyr treatment parallels significant downregulation of most acylsugar and flavonoid metabolic genes (Table 14). Additionally, putative orthologs of these transcription factor genes in *S. lycopersicum* exhibit trichome-enriched expression, similar to many genes involved in acylsugar and flavonoid metabolism (Ning et al., 2015). Together, these results suggest important roles for these transcription factors in specialized metabolic pathways of *S. pennellii* trichomes.

2.5. Methods

2.5.1. Plant Materials

Seeds of all *S. pennellii* accessions were obtained from the C. M. Rick Tomato Genetics Resource Center (University of California, Davis). Seeds were treated with a solution of 20% commercial bleach (1.2% sodium hypochlorite) for 20 minutes, and then germinated on moistened filter paper. Seedlings were transferred to soil at the cotyledon stage, and plants were grown in a growth chamber (24°C day/ 20°C night

temperature, 150 mMol m^{-2} s^{-1} photosynthetically active radiation, 16 h photoperiod, and 75% humidity).

2.5.2. RNA Sequencing (RNA-Seq) Library Preparation and Sequencing

Three biological replicates were used for each accession in the 'LOW' vs. 'HIGH' comparison, and for the 'MEDIUM' accession LA1302. Four biological replicates were used for each accession in the LA1920 vs. LA0716 comparison. Before RNA extraction, secreted acylsugars were removed from the leaf surface of 10-week-old plants by dipping them in ethanol for 2-3 seconds. Leaves were then immediately frozen in liquid nitrogen and stored at -80°C until further use. Total RNA was extracted from leaves using the RNAqueous Total RNA Isolation kit (ThermoFischer). Genomic DNA was removed using the TURBO DNA-free kit (ThermoFischer). Quality of each RNA sample was analyzed with the Agilent 2200 TapeStation software A01.04 (Agilent Technologies). RNA-Seq libraries were prepared from polyA^+-selected RNA samples using the Illumina TruSeq RNA Library Preparation kit v2, and then sequenced on the Illumina HiSeq 2500 v4 125x125-bp paired-end sequencing platform (High Output Mode) following manufacturer's specifications. Sequencing was performed at the Texas A&M Genomics and Bioinformatics service center. Sequence cluster identification, quality prefiltering, base calling and uncertainty assessment were done in real time using Illumina's HCS v2.2.68 and RTA v1.18.66.3 software with default parameter setting. Sequencer.bcl basecall files were demultiplexed and formatted into .fastq files using bcl2fastq v2.17.1.14 script conFigureBclToFastq.pl. Sequencing reads were submitted to

the National Center for Biotechnology Information (NCBI) Sequence Read Archive under accession number SRP136022.

2.5.3. Quality Control and Mapping of Sequencing Reads

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

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

Supplemental Table 2. Reads mapped to selected loci of interest were visualized with IGV (Thorvaldsdottir et al., 2013).

2.5.4. Differential Gene Expression Analyses

Uniquely-mapped reads from TopHat2 were counted for all *S. pennellii* annotated genes using HTSeq-Count (Anders et al., 2015). The following parameters were used for the counting process: -f bam, -r name, -s no, -m union, -a 20. These count files were used for identification of differentially expressed genes using edgeR (Robinson et al., 2010). Genes with very low expression levels were filtered out before differential testing. For comparison between the 'LOW' and 'HIGH' groups, genes with more than 1 count per million (CPM) in at least six samples were retained for analysis after read counts were adjusted for transcript lengths. Library sizes were recalculated after the filtering process, and all the samples had post-filter library sizes of more than 99.70%. Tagwise dispersions were calculated (prior.df= 3), and genes were identified as differentially expressed when *P* values, corrected for multiple testing using Benjamini-Hochberg multiple testing correction (Benjamini and Hochberg, 1995), were less than 0.05 (false discovery rate < 0.05), and had at least 2-fold expression differences. For comparison between LA1920 and LA0716, we used the same pipeline, except genes with more than 1 CPM in at least two samples were retained for analysis, and tagwise dispersions were calculated with prior.df= 9. In this analysis, all the samples showed more than 99.85% post-filter library sizes.

2.5.5. Functional Annotation of Differentially Expressed Genes

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

2.5.6. Identification of Putative Orthologs and Their Trichome-Enriched Expression

We used reciprocal BLAST to identify putative orthologs of *S. pennellii* genes in *S. lycopersicum*. An all-versus-all BLAST was performed between annotated proteins of *S. pennellii* v2.0 (Bolger et al., 2014a) and *S. lycopersicum* ITAG2.3 (The Tomoto Genome Consortium, 2012; Fernandez-Pozo et al., 2015) with the following parameters-

minimum percentage identity=70, minimum percentage query coverage= 50. Pairs with lowest e-value for each BLASTp comparison were selected as putative orthologs. Gene expression profiles of *S. pennellii* putative orthologs in *S. lycopersicum* were obtained from Ning et al. (2015).

2.5.7. Gene Co-Expression Network Analysis

Simple correlation analysis among expression profiles (FPKM values in our 29 samples) of selected genes was performed using the 'cor' function in R programming language (R Core Team, 2014). Gene co-expression network analysis and identification of modules (clusters of strongly correlated genes) were performed using the R package WGCNA (Langfelder and Horvath, 2008). Normalized gene expression data (FPKM values) in our 38 samples were used as the input, and the 19,378 genes which passed our filtering criteria for minimum expression levels during 'LOW' vs. 'HIGH' differential testing (*Sopen12g004230* and *Sopen12g004240* were merged) were selected for WGCNA analysis. A matrix of pairwise Spearman's rank correlation coefficients (SRCCs) between these 19,378 genes was created and then transformed into an adjacency (connection strength) matrix using 'signed' network type and a soft threshold (β) value of 12. The value of β was determined based on the scale-free topology fit index ($R^2 > 0.85$) and low mean connectivity. The weighted adjacency matrix was then converted to a topological overlap matrix (TOM), and genes were hierarchically clustered based on dissimilarity (dissTOM= 1 - TOM) of the topological overlap. Co-expression modules were defined as the branches of the clustering tree (dendrogram)

using the following parameters in ‘cutreeDynamic’ function: distM= disTOM, deepSplit=2, minClusterSize= 50. A module eigengene (ME; representative of gene expression profile in each module) was defined as the first principle component in each module. A set of 41 modules were obtained. For each gene, a ‘module membership’ was calculated as the correlation (SRCC) between expression profile of that gene and module eigengene. Associations between external traits and modules were measured as determined by SRCC between trait data and module eigengenes; ‘gene significance’ values for each gene in a module of interest were calculated as the correlation (SRCC) between expression profile of a gene and trait data. The top-50 intramodular connections for selected genes were visualized by VisANT (Hu et al., 2005).

2.5.8. Inhibitor Study (*S. pennellii* Accession LA0716)

Compound leaves of *S. pennellii* LA0716 bearing five leaflets were used for the inhibitor study. Imazapyr (Sigma-Aldrich, USA) was dissolved in distilled water, and administered as 0.1 mM and 1 mM solutions for 36 hours. Cuttings were transferred to small trays containing imazapyr or control solutions, and placed under growth chamber conditions. Total RNA was extracted as described before, and RNA-Seq libraries were prepared using Illumina TruSeq Stranded RNA LT kit. Libraries were sequenced on the Illumina HiSeq 4000 150x150 bp paired-end sequencing platform at the Texas A&M Genomics and Bioinformatics service center, College Station. Raw reads were processed with Trimmomatic (Bolger et al., 2014b), and read-mapping to *S. pennellii* genome was performed with Tophat2 (Kim et al., 2013) using the same parameters as described

before, except with the following change: -library-type= fr-firststrand. Uniquely-mapped reads from TopHat2 were counted with HTSeq-Count (Anders et al., 2015) using the following options: -f bam, -r name, -s reverse, -m union, -a 20. Differentially expressed genes (FDR < 0.05; $\log_2$FC > 1) were identified using edgeR (Robinson et al., 2010) with tagwise dispersions (prior.df= 12), and genes with more than 1 CPM in at least three samples were retained for the analysis.

2.5.9. qRT-PCR (*S. pennellii* Accession LA0716)

Total RNA from *S. pennellii* LA0716 was extracted using the RNAqueous Total RNA Isolation kit (ThermoFischer), followed by genomic DNA digestion using the TURBO DNA-free kit (ThermoFischer), and 1 µg of total RNA was reverse transcribed using the SuperScript IV VILO Master Mix (ThermoFischer), following manufacturer's guidelines. qPCR was performed using SYBR Green Master Mix (Bio-Rad) with QuantStudio 6 Flex System, and the following PCR program was used- 50°C for 2 minutes, 95°C for 10 minutes, followed by 40 cycles of (95°C for 15 seconds and 60°C for 1 minute). Differential gene expression was measured using comparative C_T ($\Delta\Delta C_T$) method. Three housekeeping genes (actin, ubiquitin, and TATA-box binding protein) were assessed for their use as endogenous control(s), and final normalization was performed with the ubiquitin gene *Sopen02g027600*. For VIGS qRT-PCR analysis, similar-sized young leaves were used for RNA extraction, and qRT-PCR primers were designed outside the VIGS-targeted region. A list of primers used in this study is given in Table 15.

Table 15. List of primers used in this study.

Gene	Purpose	Primer Sequence (5' → 3')
Sopen12g004230-Sopen12g004240	RT-PCR check (misannotation)	CTCTGTTTCTGTAGGCATTTTTTGC
		CCTTTCACGTGCTTGCTCC
Sopen02g027600 (ubiquitin)	qRT-PCR	CAAGTAGACCACCACCACTTC
		CTGATGCCGCTCCAAATGT
Sopen12g002290 (SpASAT1)	qRT-PCR	CCAAACTGTAACCATCACTAACCTTG
		GTCGCAGCTAAACTTTGTCATTTTG
Sopen08g002520 (KAS III)	VIGS	CGACGACAAGACCCTGAATATTGCAGAAGCCCATGT
		CGACGACAAGACCCTTGTGCAGGCCTGTGATATC
Sopen08g002520 (KAS III)	qRT-PCR	GTGTTCCAATACTGCTAAAGGTG
		GGTAACTCTCAGGCTTGGTAC
Sopen12g004240 (KAS IV/II-like)	VIGS	CGACGACAAGACCCTGCCACCTTAATATCCAATC
		GAGGAGAAGAGCCCTGGATTTGTAATGGGTGAAGG
Sopen12g004240 (KAS IV/II-like)	qRT-PCR	GGCTAAGTTTCAGTATTCCAAGAC
		TTGTGCACTGCCACTAAATG

2.5.10. VIGS and Metabolite Profiling (*S. pennellii* Accession LA0716)

VIGS constructs were designed using the SGN online tool (http://vigs.solgenomics.net/). Approximately 495- to 510-bp target fragments were cloned into the pTRV2-LIC vector (Dong et al., 2007). *Agrobacterium tumefaciens* strain GV3101 carrying pTRV1 and pTRV2 vectors were grown overnight at 28°C. Overnight cultures were centrifuged, and cells were resuspended in infiltration medium (10 mM MES, pH 5.5; 10 mM $MgCl_2$; 200 µM acetosyringone). TRV1 and TRV2 suspensions were incubated at room temperature for 3 hours, and then mixed in a 1:1

ratio with a final OD_{600} of 1. Infiltration of *S. pennellii* LA0716 was performed at the first true leaf stage, and all plants were grown for about six weeks in the same growth chamber mentioned before. A *phytoene desaturase* gene (*Sopen03g041530*) was used as a positive control to assess the efficiency of VIGS.

For the analysis of acylsugar acyl chain composition, secreted acylsugars were extracted from similar-sized young leaves by dipping them in ethanol. Ethanol was evaporated, and acylsugars were dissolved in *n*-heptane. Transesterification reaction was performed by adding 500 µL of 20% sodium ethoxide to 1 mL of heptane mixture at room temperature for 10 minutes with gentle vortexing. The reaction mixture was washed twice with 500 µL of saturated sodium chloride solution, and the heptane layer was used for gas chromatography-mass spectrometry (GC-MS) analysis using Trace GC Ultra with DSQII system (Thermo Scientific). A 30 m x 0.25 mm column with 0.25 µm thickness (DB-5MS; Agilent) was used as the GC column. 1 µL of heptane extract was used for split injection (1:50) with injector temperature of 225°C, and the following GC temperature program was used: starting temperature of 30°C held for 2 minutes, increase at 15°C/minute up to 300°C and held for 5 min. The carrier gas flow was 1.5 mL/minute. Thermo Xcalibur (v3.0.63) and Qual Browser applications were used for data acquisition and analysis.

2.5.11. Accession Numbers

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

3. CONCLUSIONS AND FUTURE DIRECTIONS

3.1. Transcriptomics Approaches for This Study

With the advent of 'omics' era and next-generation sequencing (NGS) technologies, there has been an explosion in the number of sequenced plant genomes and transcriptomes. For example, genomes of more than 500 plant species have been sequenced till now (http://www.ncbi.nlm.nih.gov/genome/browse/). Such an enormous amount of data will certainly help in crop improvement and plant metabolic engineering in general. However, one major challenge is how to extract biologically meaningful ideas from all of these data. With decreasing NGS cost and development of computational methods in biological sciences, large-scale biology has prospered tremendously in the last decade. These advances allowed us to use transcriptomics approaches to identify candidate genes involved in acylsugar metabolism.

3.1.1. Approach 1: Differential Gene Expression Analysis

RNA-Seq is an experimental tool for transcriptome sequencing of biological samples using NGS technologies. The primary purposes of this tool include assembly and annotation of transcriptomes and gene discovery. NGS technologies' ability to generate deep-coverage information about RNA molecules can also be used for the analysis of differential gene expression between two or more biological samples such as between different species, different varieties, different tissues, different environmental conditions, etc. Here, we used RNA-Seq tools to exploit natural variation among *S.*

pennellii accessions that produce acylsugars. In the first step, we identified genes that are differentially expressed between low- and high-acylsugar-producing accessions. The 'HIGH' vs. 'LOW' list of 1679 DEGs contained some genes that are accession-specific; for example some of the 1679 DEGs had high expression levels in only two of the three LOW accessions (LA1911 and LA1912), but no expression in other low- (or high-) acylsugar-producing accessions. Such genes were absent from the list of LA0716/LA1920 DEGs, since they are expressed in neither LA0716 nor LA1920. Subsequently, when we compared the 'HIGH' vs. 'LOW' DEGs with the LA0716 vs. LA1920 DEGs, such accession-specific DEGs were filtered out. Similarly, the list of LA0716/LA1920 DEGs (3524) includes many accession-specific DEGs. Our method of utilizing two independent comparisons filtered out those accession-specific genes and gave more robust list of DEGs (1087 genes) that are truly differentially expressed between high- and low-acylsugar-producing accessions. When we separately performed differential gene expression analysis between all low- and all high-acylsugar-producing accessions, many accession-specific genes were still present in the list of DEGs, and hence we did not adopt that method. Candidacy of acylsugar metabolic genes were further checked with their expression levels in the 'MEDIUM' accession LA1302, since known genes showed intermediate expression levels in LA1302. Additionally, putative tomato orthologs of many acylsugar candidate DEGs are preferentially expressed in isolated trichomes. Because acylsugars are secreted by trichome tip cells (Fobes et al., 1985), it is reasonable to expect that genes involved in acylsugar secretion will exhibit trichome-specific/ trichome-enriched expression pattern.

3.1.2. Approach 2: A Chemical Genetics Approach

Plants can synthesize each of the 20 standard amino acids *de novo*. Branched-chain amino acids are essential amino acids, which are not synthesized by humans and other animals. Plant-synthesized BCAAs are the main sources of these nutrients in human diets. BCAA metabolic pathway has been studied in Arabidopsis and other plants, partly because BCAA biosynthesis is an ideal target for selective herbicides. The sulfonylurea and imidazolinone herbicides are the two most common herbicides that inhibit BCAA biosynthesis. These herbicides target the enzyme ALS, which catalyzes the committing step in the BCAA biosynthetic pathway. We used imazapyr at two different concentrations to inhibit the enzyme ALS, and subsequently BCAA biosynthesis. Results showed that known and putative high-confidence acylsugar candidate genes are downregulated in response to imazapyr in a concentration-dependent manner. This approach significantly narrowed the list of initial 1087 DEGs to 171 DEGs. RNA-Seq data from our experiments will be useful for future genetic studies involving BCAA metabolism in *S. pennellii* and other plants.

3.1.3. Approach 3: Gene Co-Expression Network Analysis

A network analysis is useful for identifying missing genes in a pathway and also for gene function discovery. For gene co-expression network analysis, sample size is a critical factor. Like other biological experiments, the larger the sample size, the more robust and refined network inferences are since the statistical power is increased when

there is more biological variation (Li et al., 2015). Due to limitations in cost, time, biological material availability and other factors, many gene co-expression network analyses are performed with small sample sizes. However, for WGCNA, at least 20 biological samples are recommended (Langfelder and Horvath, 2008). We used 38 samples for WGCNA and robust results were obtained. This third approach identified 57 genes that responded to imazapyr and are strongly co-expressed with known acylsugar biosynthetic genes (ASATs and the invertase gene).

3.2. Acylsugar Metabolic Genes

In animals and plants (primary metabolism), BCAAs are used to generate energy and/or TCA cycle intermediates via multiple catabolic steps (Binder, 2010). Unlike animals, plants are capable of both *de novo* synthesis and degradation of BCAAs. These two counteracting processes therefore must be regulated carefully to maintain a balanced homeostasis for these biomolecules. In *Arabidopsis thaliana* and other plants, most of the enzymes involved in BCAA breakdown have been located in mitochondria and/or peroxisome (Mooney et al., 2002; Taylor et al., 2004; Reumann et al., 2009). Therefore, compartmentalization physically separates BCAA catabolism in mitochondria-peroxisome from BCAA anabolism in chloroplasts. Genes involved in both BCAA biosynthesis in chloroplasts and BCAA to BCFA-CoA conversion in mitochondria were upregulated in high-acylsugar-producing accessions, which validated our approach1. Sequence similarity between *Sopen12g004240*'s KAS domain and *Cuphea* KAS IV/KAS II-like enzymes, which synthesize medium-chain (C8-C12) SCFAs, suggested a

role of *Sopen12g004240* (in collaboration with *Sopen08g002520*; KAS III) in acylsugar SCFA synthesis. VIGS of these two candidate genes confirmed their role in SCFA biosynthesis.

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

Putative peroxisomal locations for the four AAE1 members (Reumann et al., 2009) are consistent with their strong connection with *Sopen11g007710*, which is predicted to encode an Mpv17/PMP22 family peroxisomal membrane protein. It should be noted that PMP22 is a major component of peroxisomal membranes, and Mpv17 is a closely related protein that may have a mitochondrial location (https://pfam.xfam.org/family/PF04117). However, subcellular localization prediction tool (http://www.jci-bioinfo.cn/pLoc-mPlant/) indicates that our *Sopen11g007710* (annotated as Mpv17 / PMP22 family) has a putative peroxisomal location, similar to the predicted peroxisomal locations of the AAE1 members. PMP22 is presumably involved

in controlling permeability of peroxisomal membrane (Brosius et al., 2002), and *Sopen11g007710* showed strong co-expression with known and candidate acylsugar metabolic genes. These results suggest an important role of peroxisome in acylsugar SCFA metabolism.

It is not known how acylsugars are secreted out of trichome cells, or whether acylglucose and acylsucrose molecules are exported by the same mechanism. The SCPL GAT, which was reported to catalyze the second step in phase 2 of acylglucose biosynthesis, has a cleavable N-terminal 18-aa signal peptide, and acylsugars (or at least acylglucose molecules) had been suggested to be secreted via a secretory vesicle pathway (Li et al., 1999; Li and Steffens, 2000). However, to our knowledge, there is no direct evidence to support this hypothesis, and recent work strongly contradicts the involvement of the SCPL GAT in acylglucose biosynthesis (Leong et al., 2019). Here, we identified five ABC transporter DEGs, and three of them were upregulated in high-acylsugar-producing accessions. Putative tomato orthologs of these three upregulated DEGs (*Sopen12g034820*, *Sopen04g005380*, and *Sopen03g001870*) show 143-, 160-, and 470-fold, respectively higher expression in isolated trichomes than in underlying tissues, with expression of *Solyc03g005860* being very similar to that of trichome-specific ASAT genes (Ning et al., 2015). These three candidate genes also show strong intramodular connectivity with ASAT and acylsugar phase 1 genes. On the other hand, putative *S. lycopersicum* orthologs of the other ABC transporter DEGs (*Sopen04g023150* and *Sopen01g047950*), which were downregulated in high-acylsugar-producing accessions, showed similar expression levels in both trichomes and

underlying tissues (Ning et al., 2015), and not downregulated in response to imazapyr (Table 12). It will be interesting to see if more than one ABC transporter are involved in acylsugar export (e.g. acylsuglucose and acylsucrose-specific ABC transporters).

3.3. Transcription Factors

Transcription factors regulate gene expression in response to different internal and external signals. We identified some transcription factor genes that responded to imazapyr, and acylsugar metabolic genes were significantly inhibited, suggesting a possible role of these transcription factors in the regulation of acylsugar biosynthesis. Study of transcription factors and their target sites (transcription factor-promoter interaction) can be used to construct a gene regulatory network. If information is available about transcription factors and their target genes in a particular cell/tissue type at a specific time point, such information can build a transcription factor regulatory network, in which nodes represent transcription factors and their target genes, whereas directed edges represent target gene regulation by transcription factors. Using information from ChIP-chip and biochemical studies, such a regulatory network was constructed in *Saccharomyces cerevisiae*, and it was concluded that regulatory network exhibits a multi-tiered, or hierarchical structure (Jothi et al., 2009). Transcription factors in the regulatory network were classified into three hierarchical layers (top, core, and bottom). The top-layer transcription factors are relatively abundant and long-lived at protein level, and exhibit a higher degree of variability in expression (noisy) compared with the core- and bottom-layer transcription factors. The bottom-layer components

showed stricter regulation of expressional states and very low levels of noise. Higher degrees of variability in expression of top-layer transcription factors may provide greater adaptability, whereas the bottom-layer transcription factors may have a buffering role to minimize noise propagation and ensure fidelity in regulation (Jothi et al., 2009; Sinha et al., 2016). Transcription factor genes that showed robust co-expression with known and candidate acylsugar metabolic genes may constitute the bottom-layer in the regulatory cascade, since their expressions were strictly regulated in response to the imazapyr treatment. On the other hand, the top-layer transcription factors that can potentially regulate the bottom-layer transcription factors would not exhibit such a robust response to imazapyr and would not be captured from our analysis. A global gene regulatory network needs to be constructed in order to capture these transcription factors.

3.4. Future Directions

In recent years, considerable progress has been made identifying genes involved in phase 2 of acylsugar biosynthesis in different solanaceous plants (Schilmiller et al., 2012; Schilmiller et al., 2015; Fan et al., 2016; Schilmiller et al., 2016; Moghe et al., 2017; Nadakuduti et al., 2017; Leong et al., 2019). However, genes involved in several other aspects of acylsugar metabolism have remained unidentified. Here, RNA-Seq tools were used to identify these missing/unknown genes. Our results suggest that a majority of the acylsugar metabolic pathway, from acyl chain synthesis to acylsugar secretion, is under a remarkable common regulatory mechanism. Although the roles we have proposed for the candidate genes remain to be confirmed by genetic or biochemical

studies. Confirmation of common regulatory factors, such as the AP2-family transcription factors, will be useful in genetic manipulation of acylsugar production. Our work also revealed differential expression of two defense networks among *S. pennellii* accessions, and raises interesting questions about evolution of acylsugar pathway and defense pathways in this wild tomato species. Overall, data presented here should serve as a valuable resource for increasing our understanding of acylsugar metabolism and defense in *S. pennellii* and other plants.

Future work mostly involves functional validation of candidate genes. Before that, the remaining candidate genes can be further filtered based on their trichome-enriched or trichome-specific expression profile. RNA samples from isolated stem trichomes and shaved underlying tissues will be used to quantify transcript levels of the remaining candidate genes. VIGS can be applied for functional validation of high-priority candidate genes. However, one major drawback is the variable and incomplete nature of VIGS in *S. pennellii*. To overcome this issue, a modified TRV-based vector has been developed (Tian et al., 2014). The green fluorescent protein (GFP) was fused to the 3' end of TRV2 coat protein. Presence of TRV-GFP could be monitored by using a hand-held UV lamp, and no significant decrease in infection efficiency could be observed. Similarly, acylsugar candidate genes can be cloned in such a TRV-GFP vector, and metabolite profiling can be performed from selective area-specific silenced regions. Other challenges in functional validation include availability of biochemical assay. Often, lack of isolated or synthesized substrate can be a potential limitation.

Additionally, if a biological function is carried out by more than one protein, gene silencing of one gene may not exhibit desired phenotype.

Transgenic plants have been used extensively to improve agronomic traits. The *S. pennellii* acylsugar biosynthetic pathway is very short (only 3-4 genes in phase 2) and therefore has the potential to be transferred into other plants with well-known agronomic features such as tomato, potato, and tobacco (*Nicotiana tabaccum*). These plants belong to the same family (Solanaceae) as *S. pennellii*, and similar types of acylsugars are also produced at low levels in these plants (Kroumova et al., 2016). Structural and regulatory genes can be expressed under the control of their native promoters in these plants to see if higher levels of acylsugar production can be achieved in these plants. Candidate genes can also be expressed under the control of a strong, constitutive viral promoter such as the CaMV *35S* promoter; however, expression under such a strong promoter may have deleterious consequences if excessive production of acylsugars exhausts the plant's carbon reserve.

Well-known plant transformation protocols are available for some solanaceous plants, including tomato and tobacco. Candidate genes with their native promoter can be cloned into a plant transformation vector and then introduced into other solanaceous plants via *Agrobacterium tumefaciens*-mediated transformation. For example, open reading frame region along with its promoter region can be easily cloned into a binary vector such as pCAMBIA 3300, and tobacco leaf discs (*Nicotiana tabaccum* cv. SR1) can be transformed by co-cultivation with *Agrobacterium tumefaciens* (either strain LBA4404 or strain GV3101 harboring recombinant vector). Leaf discs with successful

transformation events (selected with a marker, such as kanamycin or phosphinothricin) can be cultured under *in vitro* conditions on Murashige and Skoog basal medium supplemented with necessary phytohormones and carbon source. Properly differentiated plants can be transferred to soil and grown for further studies. Production of acylsugars in transgenic plants can be evaluated following methods described earlier, and it can be checked if transgenic plants can be provided with better defense against biotic agents.

REFERENCES

Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Cech, M., Chilton, J., Clements, D., Coraor, N., Eberhard, C., Gruning, B., Guerler, A., Hillman-Jackson, J., Von Kuster, G., Rasche, E., Soranzo, N., Turaga, N., Taylor, J., Nekrutenko, A., and Goecks, J. (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res. **44**: W3-W10.

Alba, J.M., Montserrat, M., and Fernandez-Munoz, R. (2009). Resistance to the two-spotted spider mite (*Tetranychus urticae*) by acylsucroses of wild tomato (*Solanum pimpinellifolium*) trichomes studied in a recombinant inbred line population. Exp. Appl. Acarol. **47**: 35-47.

Albinsky, D., Sawada, Y., Kuwahara, A., Nagano, M., Hirai, A., Saito, K., and Hirai, M.Y. (2010). Widely targeted metabolomics and coexpression analysis as tools to identify genes involved in the side-chain elongation steps of aliphatic glucosinolate biosynthesis. Amino Acids **39**: 1067-1075.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq-a Python framework to work with high-throughput sequencing data. Bioinformatics **31**: 166-169.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E.,

Ringwald, M., Rubin, G.M., and Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The gene ontology consortium. Nat. Genet. **25**: 25-29.

Atwell, S., Huang, Y.S., Vilhjalmsson, B.J., Willems, G., Horton, M., Li, Y., Meng, D., Platt, A., Tarone, A.M., Hu, T.T., Jiang, R., Muliyati, N.W., Zhang, X., Amer, M.A., Baxter, I., Brachi, B., Chory, J., Dean, C., Debieu, M., de Meaux, J., et al. (2010). Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. Nature **465**: 627-631.

Bachan, S., and Dinesh-Kumar, S.P. (2012). Tobacco rattle virus (TRV)-based virus-induced gene silencing. Methods Mol. Biol. **894**: 83-92.

Balandrin, M.F., Klocke, J.A., Wurtele, E.S., and Bollinger, W.H. (1985). Natural plant chemicals: sources of industrial and medicinal materials. Science **228**: 1154-1160.

Balcke, G.U., Bennewitz, S., Bergau, N., Athmer, B., Henning, A., Majovsky, P., Jiménez-Gómez, J.M., Hoehenwarter, W., and Tissier, A. (2017). Multi-omics of tomato glandular trichomes reveals distinct features of central carbon metabolism supporting high productivity of specialized metabolites. Plant Cell **29**: 960-983.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Series B Stat. Methodol. **57**: 289-300.

Binder, S. (2010). Branched-chain amino acid metabolism in *Arabidopsis thaliana*. Arabidopsis Book **8**: e0137.

Blauth, S.L., Steffens, J.C., Churchill, G.A., and Mutschler, M.A. (1999). Identification of QTLs controlling acylsugar fatty acid composition in an intraspecific population of *Lycopersicon pennellii* (Corr.) D'Arcy. Theor. Appl. Genet. **99**: 373-381.

Bolger, A., Scossa, F., Bolger, M.E., Lanz, C., Maumus, F., Tohge, T., Quesneville, H., Alseekh, S., Sorensen, I., Lichtenstein, G., Fich, E.A., Conte, M., Keller, H., Schneeberger, K., Schwacke, R., Ofner, I., Vrebalov, J., Xu, Y., Osorio, S., Aflitos, S.A., et al. (2014a). The genome of the stress-tolerant wild tomato species *Solanum pennellii*. Nat. Genet. **46**: 1034-1038.

Bolger, A.M., Lohse, M., and Usadel, B. (2014b). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30**: 2114-2120.

Bonierbale, M.W., Plaisted, R.L., Pineda, O., and Tanksley, S.D. (1994). Qtl analysis of trichome-mediated insect resistance in potato. Theor. Appl. Genet. **87**: 973-987.

Brosius, U., Dehmel, T., and Gartner, J. (2002). Two different targeting signals direct human peroxisomal membrane protein 22 to peroxisomes. J. Biol. Chem. **277**: 774-784.

Cahais, V., Gayral, P., Tsagkogeorga, G., Melo-Ferreira, J., Ballenghien, M., Weinert, L., Chiari, Y., Belkhir, K., Ranwez, V., and Galtier, N. (2012). Reference-free transcriptome assembly in non-model animals from next-generation sequencing data. Mol. Ecol. Resour. **12**: 834-845.

Cao, J., Schneeberger, K., Ossowski, S., Gunther, T., Bender, S., Fitz, J., Koenig, D., Lanz, C., Stegle, O., Lippert, C., Wang, X., Ott, F., Muller, J., Alonso-Blanco, C., Borgwardt, K., Schmid, K.J., and Weigel, D. (2011). Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. Nat. Genet. **43**: 956-963.

Chan, E.K., Rowe, H.C., and Kliebenstein, D.J. (2010). Understanding the evolution of defense metabolites in *Arabidopsis thaliana* using genome-wide association mapping. Genetics **185**: 991-1007.

Chen, F., Tholl, D., D'Auria, J.C., Farooq, A., Pichersky, E., and Gershenzon, J. (2003). Biosynthesis and emission of terpenoid volatiles from Arabidopsis flowers. Plant Cell **15**: 481-494.

Chen, Y.S., Lun, A.T.L., and Smyth, G.K. (2014). Differential expression analysis of complex RNA-seq experiments using edgeR. Front. Probab. Stat. Sc. 8: 51-74.

Chortyk, O.T., Kays, S.J., and Teng, Q. (1997). Characterization of insecticidal sugar esters of *Petunia*. J. Agric. Food Chem. **45**: 270-275.

Chortyk, O.T., Severson, R.F., Cutler, H.C., and Sisson, V.A. (1993). Antibiotic activities of sugar esters isolated from selected *Nicotiana* species. Biosci. Biotech. Bioch. **57**: 1355-1356.

Cona, A., Rea, G., Angelini, R., Federico, R., and Tavladoraki, P. (2006). Functions of amine oxidases in plant development and defence. Trends Plant Sci. **11**: 80-88.

Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics **21**: 3674-3676.

Consortium, T.T.G. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature **485**: 635-641.

Couto, D., and Zipfel, C. (2016). Regulation of pattern recognition receptor signalling in plants. Nat. Rev. Immunol. **16**: 537-552.

de Kraker, J.W., Luck, K., Textor, S., Tokuhisa, J.G., and Gershenzon, J. (2007). Two Arabidopsis genes (IPMS1 and IPMS2) encode isopropylmalate synthase, the branchpoint step in the biosynthesis of leucine. Plant Physiol. **143**: 970-986.

Dehesh, K., Edwards, P., Fillatti, J., Slabaugh, M., and Byrne, J. (1998). KAS IV: a 3-ketoacyl-ACP synthase from *Cuphea* sp. is a medium chain specific condensing enzyme. Plant J. **15**: 383-390.

Dixon, R.A., and Strack, D. (2003). Phytochemistry meets genome analysis, and beyond. Phytochemistry **62**: 815-816.

Dong, Y., Burch-Smith, T.M., Liu, Y., Mamillapalli, P., and Dinesh-Kumar, S.P. (2007). A ligation-independent cloning tobacco rattle virus vector for high-throughput virus-induced gene silencing identifies roles for *NbMADS4-1* and *-2* in floral development. Plant Physiol. **145**: 1161-1170.

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

Fan, P., Miller, A.M., Schilmiller, A.L., Liu, X., Ofner, I., Jones, A.D., Zamir, D., and Last, R.L. (2016). *In vitro* reconstruction and analysis of evolutionary variation of the tomato acylsucrose metabolic network. Proc. Natl. Acad. Sci. USA **113**: E239-248.

Fernandez-Pozo, N., Menda, N., Edwards, J.D., Saha, S., Tecle, I.Y., Strickler, S.R., Bombarely, A., Fisher-York, T., Pujar, A., Foerster, H., Yan, A., and Mueller, L.A. (2015). The Sol Genomics Network (SGN)-from genotype to phenotype to breeding. Nucleic Acids Res. **43**: D1036-1041.

Figueiredo, A., Monteiro, F., and Sebastiana, M. (2014). Subtilisin-like proteases in plant-pathogen recognition and immune priming: a perspective. Front. Plant Sci. **5**: 739.

Fobes, J.F., Mudd, J.B., and Marsden, M.P. (1985). Epicuticular lipid accumulation on the leaves of *Lycopersicon pennellii* (Corr.) D'Arcy and *Lycopersicon esculentum* Mill. Plant Physiol. **77**: 567-570.

Galili, G., Amir, R., and Fernie, A.R. (2016). The regulation of essential amino acid synthesis and accumulation in plants. Annu. Rev. Plant Biol. **67**: 153-78.

Garcia, M.D., Nouwens, A., Lonhienne, T.G., and Guddat, L.W. (2017). Comprehensive understanding of acetohydroxyacid synthase inhibition by different herbicide families. Proc. Natl. Acad. Sci. USA **114**: E1091-E1100.

Gentile, A.G., Webb, R.E., and Stoner, A.K. (1968). Resistance in *Lycopersicon* and *Solanum* to greenhouse whiteflies. J. Econ. Entomol. **61**: 1355-57.

Geu-Flores, F., Sherden, N.H., Courdavault, V., Burlat, V., Glenn, W.S., Wu, C., Nims, E., Cui, Y., and O'Connor, S.E. (2012). An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis. Nature **492**: 138-142.

Ghangas, G.S., and Steffens, J.C. (1993). UDPglucose: fatty acid transglucosylation and transacylation in triacylglucose biosynthesis. Proc. Natl. Acad. Sci. USA **90**: 9911-9915.

Giardine, B., Riemer, C., Hardison, R.C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J., Miller, W., Kent, W.J., and Nekrutenko, A. (2005). Galaxy: A platform for interactive large-scale genome analysis. Genome Res. **15**: 1451-1455.

Giddings, L.A., Liscombe, D.K., Hamilton, J.P., Childs, K.L., DellaPenna, D., Buell, C.R., and O'Connor, S.E. (2011). A stereoselective hydroxylation step of alkaloid biosynthesis by a unique cytochrome P450 in *Catharanthus roseus*. J. Biol. Chem. **286**: 16751-16757.

Gigolashvili, T., Engqvist, M., Yatusevich, R., Muller, C., and Flugge, U.I. (2008). HAG2/MYB76 and HAG3/MYB29 exert a specific and coordinated control on the regulation of aliphatic glucosinolate biosynthesis in *Arabidopsis thaliana*. New Phytol. **177**: 627-642.

Gigolashvili, T., Berger, B., Mock, H.P., Muller, C., Weisshaar, B., and Flugge, U.I. (2007). The transcription factor HIG1/MYB51 regulates indolic glucosinolate biosynthesis in *Arabidopsis thaliana*. Plant J. **50**: 886-901.

Goffreda, J.C., Mutschler, M.A., Ave, D.A., Tingey, W.M., and Steffens, J.C. (1989). Aphid deterrence by glucose esters in glandular trichome exudate of the wild tomato *Lycopersicon pennellii*. J. Chem. Ecol .**15**: 2135-2147.

Gongora-Castillo, E., Childs, K.L., Fedewa, G., Hamilton, J.P., Liscombe, D.K., Magallanes-Lundback, M., Mandadi, K.K., Nims, E., Runguphan, W., Vaillancourt, B., Varbanova-Herde, M., Dellapenna, D., McKnight, T.D., O'Connor, S., and Buell, C.R. (2012). Development of transcriptomic resources for interrogating the biosynthesis of monoterpene indole alkaloids in medicinal plant species. PLoS One **7**: e52506.

Gotz, S., Garcia-Gomez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talon, M., Dopazo, J., and Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. **36**: 3420-3435.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. **29**: 644-652.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., MacManes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T.,

Dewey, C.N., Henschel, R., et al. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. **8**: 1494-1512.

Halitschke, R., and Baldwin, I.T. (2003). Antisense LOX expression increases herbivore performance by decreasing defense responses and inhibiting growth-related transcriptional reorganization in *Nicotiana attenuata*. Plant J. **36**: 794.

Halitschke, R., Gase, K., Hui, D. Q., Schmidt, D. D., and Baldwin, I. T. (2003). Molecular interactions between the specialist herbivore *Manduca sexta* (Lepidoptera, Sphingidae) and its natural host *Nicotiana attenuata*. Plant Physiol. **131**: 1894-1902.

Hare, J.D. (2005). Biological activity of acyl glucose esters from *Datura wrightii* glandular trichomes against three native insect herbivores. J. Chem. Ecol. **31**: 1475-1491.

Hawthorne, D.J., Shapiro, J.A., Tingey, W.M., and Mutschler, M.A. (1992). Trichome-borne and artificially applied acylsugars of wild tomato deter feeding and oviposition of the leafminer *Liriomyza trifolii*. Entomol. Exp. Appl. **65**: 65-73.

He, Y., Mawhinney, T.P., Preuss, M.L., Schroeder, A.C., Chen, B., Abraham, L., Jez, J.M., and Chen, S. (2009). A redox-active isopropylmalate dehydrogenase functions in the biosynthesis of glucosinolates and leucine in Arabidopsis. Plant J. **60**: 679-690.

Hill, K., and Rhode, O. (1999). Sugar-based surfactants for consumer products and technical applications. Fett-Lipid **101**: 25-33.

Hirai, M.Y., Sugiyama, K., Sawada, Y., Tohge, T., Obayashi, T., Suzuki, A., Araki, R., Sakurai, N., Suzuki, H., Aoki, K., Goda, H., Nishizawa, O.I., Shibata, D., and Saito, K. (2007). Omics-based identification of Arabidopsis Myb transcription factors regulating aliphatic glucosinolate biosynthesis. Proc. Natl. Acad. Sci. USA **104**: 6478-6483.

Hu, Z., Mellor, J., Wu, J., Yamada, T., Holloway, D., and Delisi, C. (2005). VisANT: data-integrating visual framework for biological networks and modules. Nucleic Acids Res. **33**: W352-357.

Iseli, C., Jongeneel, C.V., and Bucher, P. (1999). ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. Proc. Int. Conf. Intell. Syst. Mol. Biol.: 138-148.

Jackson, P.A., Galinha, C.I., Pereira, C.S., Fortunato, A., Soares, N.C., Amancio, S.B., and Pinto Ricardo, C.P. (2001). Rapid deposition of extensin during the elicitation of grapevine callus cultures is specifically catalyzed by a 40-kilodalton peroxidase. Plant Physiol. **127**: 1065-1076.

Juvik, J.A., Berlinger, M.J., Bendavid, T., and Rudich, J. (1982). Resistance among accessions of the genera *Lycopersicon* and *Solanum* to 4 of the main insect pests of tomato in Israel. Phytoparasitica **10**: 145-156.

Juvik, J.A., Shapiro, J.A., Young, T.E., and Mutschler, M.A. (1994). Acylglucoses from wild tomatoes alter behavior and reduce growth and survival of

Helicoverpa zea and *Spodoptera exigua* (Lepidoptera, Noctuidae). J. Econ. Entomol. **87**: 482-492.

Kandra, L., and Wagner, G.J. (1990). Chlorsulfuron modifies biosynthesis of acyl acid substituents of sucrose esters secreted by tobacco trichomes. Plant Physiol. **94**: 906-912.

Kang, J., Park, J., Choi, H., Burla, B., Kretzschmar, T., Lee, Y., and Martinoia, E. (2011). Plant ABC Transporters. Arabidopsis Book **9**: e0153.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. **14**.

Kim, J., Matsuba, Y., Ning, J., Schilmiller, A.L., Hammar, D., Jones, A.D., Pichersky, E., and Last, R.L. (2014). Analysis of natural and induced variation in tomato glandular trichome flavonoids identifies a gene not present in the reference genome. Plant Cell **26**: 3272-3285.

Kliebenstein, D.J., Lambrix, V.M., Reichelt, M., Gershenzon, J., and Mitchell-Olds, T. (2001a). Gene duplication in the diversification of secondary metabolism: tandem 2-oxoglutarate-dependent dioxygenases control glucosinolate biosynthesis in Arabidopsis. Plant Cell **13**: 681-693.

Kliebenstein, D.J., Kroymann, J., Brown, P., Figuth, A., Pedersen, D., Gershenzon, J., and Mitchell-Olds, T. (2001b). Genetic control of natural variation in Arabidopsis glucosinolate accumulation. Plant Physiol. **126**: 811-825.

Knill, T., Reichelt, M., Paetz, C., Gershenzon, J., and Binder, S. (2009). Arabidopsis thaliana encodes a bacterial-type heterodimeric isopropylmalate isomerase involved in both Leu biosynthesis and the Met chain elongation pathway of glucosinolate formation. Plant Mol. Biol. **71**: 227-239.

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

Kroumova, A.B., and Wagner, G.J. (2003). Different elongation pathways in the biosynthesis of acyl groups of trichome exudate sugar esters from various solanaceous plants. Planta **216**: 1013-1021.

Kroumova, A.B., Xie, Z., and Wagner, G.J. (1994). A pathway for the biosynthesis of straight and branched, odd- and even-length, medium-chain fatty acids in plants. Proc. Natl. Acad. Sci. USA **91**: 11437-11441.

Kroumova, A.B., Zaitlin, D., and Wagner, G.J. (2016). Natural variability in acyl moieties of sugar esters produced by certain tobacco and other Solanaceae species. Phytochem. **130**: 218-227.

Kroymann, J. (2011). Natural diversity and adaptation in plant secondary metabolism. Curr. Opin. Plant. Biol. **14**: 246-251.

Kuai, J.P., Ghangas, G.S., and Steffens, J.C. (1997). Regulation of triacylglucose fatty acid composition (uridine diphosphate glucose:fatty acid glucosyltransferases with overlapping chain-length specificity). Plant Physiol. **115**: 1581-1587.

La Camera, S., Gouzerh, G., Dhondt, S., Hoffmann, L., Fritig, B., Legrand, M., and Heitz, T. (2004). Metabolic reprogramming in plant innate immunity: the contributions of phenylpropanoid and oxylipin pathways. Immunol. Rev. **198**: 267-284.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics **9**: 559.

Lau, W., and Sattely, E.S. (2015). Six enzymes from mayapple that complete the biosynthetic pathway to the etoposide aglycone. Science **349**: 1224-1228.

Lawson, D.M., Lunde, C.F., and Mutschler, M.A. (1997). Marker-assisted transfer of acylsugar-mediated pest resistance from the wild tomato, *Lycopersicon pennellii*, to the cultivated tomato, *Lycopersicon esculentum*. Mol. Breeding **3**: 307-317.

Leckie, B.M., De Jong, D.M., and Mutschler, M.A. (2012). Quantitative trait loci increasing acylsugars in tomato breeding lines and their impacts on silverleaf whiteflies. Mol. Breeding **30**: 1621-1634.

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

Li, A.X., and Steffens, J.C. (2000). An acyltransferase catalyzing the formation of diacylglucose is a serine carboxypeptidase-like protein. Proc. Natl. Acad. Sci. USA **97**: 6902-6907.

Li, A.X., Eannetta, N., Ghangas, G.S., and Steffens, J.C. (1999). Glucose polyester biosynthesis. Purification and characterization of a glucose acyltransferase. Plant Physiol. **121**: 453-460.

Li, X., Svedin, E., Mo, H., Atwell, S., Dilkes, B.P., and Chapple, C. (2014). Exploiting natural variation of secondary metabolism identifies a gene controlling the glycosylation diversity of dihydroxybenzoic acids in *Arabidopsis thaliana*. Genetics **198**: 1267-1276.

Li, Y., Pearl, S.A., and Jackson, S.A. (2015). Gene networks in plant biology: Approaches in reconstruction and analysis. Trends. Plant Sci. **20**: 664-675.

Li, Y., Huang, Y., Bergelson, J., Nordborg, M., and Borevitz, J.O. (2010). Association mapping of local climate-sensitive quantitative trait loci in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA **107**: 21199-21204.

Liedl, B.E., Lawson, D.M., White, K.K., Shapiro, J.A., Cohen, D.E., Carson, W.G., Trumble, J.T., and Mutschler, M.A. (1995). Acylsugars of wild tomato *Lycopersicon pennellii* alters settling and reduces oviposition of *Bemisia argentifolii* (Homoptera, Aleyrodidae). J. Econ. Entomol. **88**: 742-748.

Liu, E., and Page, J.E. (2008). Optimized cDNA libraries for virus-induced gene silencing (VIGS) using tobacco rattle virus. Plant Methods **4**: 5.

Lu, R., Martin-Hernandez, A.M., Peart, J.R., Malcuit, I., and Baulcombe, D.C. (2003). Virus-induced gene silencing in plants. Methods **30**: 296-303.

Luu, V.T., Weinhold, A., Ullah, C., Dressel, S., Schoettner, M., Gase, K., Gaquerel, E., Xu, S., and Baldwin, I.T. (2017). O-acyl sugars protect a wild tobacco from

both native fungal pathogens and a specialist herbivore. Plant Physiol. **174**: 370-386.

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

Manabe, Y., Tinker, N., Colville, A., and Miki, B. (2007). CSR1, the sole target of imidazolinone herbicide in *Arabidopsis thaliana*. Plant Cell Physiol. **48**: 1340-1358.

Martin, J.A., and Wang, Z. (2011). Next-generation transcriptome assembly. Nat. Rev. Genet. 12: 671-682.

McDowell, E.T., Kapteyn, J., Schmidt, A., Li, C., Kang, J.H., Descour, A., Shi, F., Larson, M., Schilmiller, A., An, L., Jones, A.D., Pichersky, E., Soderlund, C.A., and Gang, D.R. (2011). Comparative functional genomic analysis of *Solanum* glandular trichome types. Plant Physiol. **155**: 524-539.

Modolo, L.V., Reichert, A.I., and Dixon, R.A. (2009). Introduction to the different classes of biosynthetic enzymes. In plant-derived natural products: Synthesis, function, and application, A.E. Osbourn and V. Lanzotti, eds (New York, NY: Springer US), pp. 143-163.

Moghe, G.D., and Last, R.L. (2015). Something old, something new: Conserved enzymes and the evolution of novelty in plant specialized metabolism. Plant Physiol. **169**: 1512-1523.

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

Mooney, B.P., Miernyk, J.A., and Randall, D.D. (2002). The complex fate of alpha-ketoacids. Annu. Rev. Plant Biol. **53**: 357-375.

Mueller, S., Hilbert, B., Dueckershoff, K., Roitsch, T., Krischke, M., Mueller, M.J., and Berger, S. (2008). General detoxification and stress responses are mediated by oxidized lipids through TGA transcription factors in Arabidopsis. Plant Cell **20**: 768-785.

Nadakuduti, S.S., Uebler, J.B., Liu, X., Jones, A.D., and Barry, C.S. (2017). Characterization of trichome-expressed BAHD acyltransferases in *Petunia axillaris* reveals distinct acylsugar assembly mechanisms within the Solanaceae. Plant Physiol. **175**: 36-50.

Niebel, A., De Almeida Engler, J., Tire, C., Engler, G., Van Montagu, M., and Gheysen, G. (1993). Induction patterns of an extensin gene in tobacco upon nematode infection. Plant Cell **5**: 1697-1710.

Ning, J., Moghe, G.D., Leong, B., Kim, J., Ofner, I., Wang, Z., Adams, C., Jones, A.D., Zamir, D., and Last, R.L. (2015). A feedback-insensitive isopropylmalate synthase affects acylsugar composition in cultivated and wild tomato. Plant Physiol. **169**: 1821-1835.

Noguchi, A., Sasaki, N., Nakao, M., Fukami, H., Takahashi, S., Nishino, T., and Nakayama, T. (2008). cDNA cloning of glycosyltransferases from Chinese

wolfberry (*Lycium barbarum* L.) fruits and enzymatic synthesis of a catechin glucoside using a recombinant enzyme (UGT73A10). J. Mol. Catal. B Enzym. **55**: 84-92.

Nordborg, M., Hu, T.T., Ishino, Y., Jhaveri, J., Toomajian, C., Zheng, H., Bakker, E., Calabrese, P., Gladstone, J., Goyal, R., Jakobsson, M., Kim, S., Morozov, Y., Padhukasahasram, B., Plagnol, V., Rosenberg, N.A., Shah, C., Wall, J.D., Wang, J., Zhao, K., et al. (2005). The pattern of polymorphism in *Arabidopsis thaliana*. PLoS Biol. **3**: e196.

Owen, C., Patron, N.J., Huang, A., and Osbourn, A. (2017). Harnessing plant metabolic diversity. Curr. Opin. Chem. Biol. **40**: 24-30.

Parra, G., Bradnam, K., and Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics **23**: 1061-1067.

Peng, C., Uygun, S., Shiu, S.H., and Last, R.L. (2015). The impact of the branched-chain ketoacid dehydrogenase complex on amino acid homeostasis in Arabidopsis. Plant Physiol. **169**: 1807-1820.

Perez-Castorena, A.L., Martinez, M., and Maldonado, E. (2010). Labdanes and sucrose esters from *Physalis sordida*. J. Nat. Prod. **73**: 1271-1276.

Pichersky, E., and Lewinsohn, E. (2011). Convergent evolution in plant specialized metabolism. Annu. Rev. Plant Biol. **62**: 549-566.

Prost, I., Dhondt, S., Rothe, G., Vicente, J., Rodriguez, M.J., Kift, N., Carbonne, F., Griffiths, G., Esquerre-Tugaye, M.T., Rosahl, S., Castresana, C., Hamberg, M., and Fournier, J. (2005). Evaluation of the antimicrobial activities of plant

oxylipins supports their involvement in defense against pathogens. Plant Physiol. **139**: 1902-1913.

Puterka, G.J., Farone, W., Palmer, T., and Barrington, A. (2003). Structure-function relationships affecting the insecticidal and miticidal activity of sugar esters. J. Econ. Entomol. **96**: 636-644.

Reumann, S., Quan, S., Aung, K., Yang, P., Manandhar-Shrestha, K., Holbrook, D., Linka, N., Switzenberg, R., Wilkerson, C.G., Weber, A.P., Olsen, L.J., and Hu, J. (2009). In-depth proteome analysis of Arabidopsis leaf peroxisomes combined with *in vivo* subcellular targeting verification indicates novel metabolic and regulatory functions of peroxisomes. Plant Physiol. **150**: 125-143.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26**: 139-140.

Rodriguez, A.E., Tingey, W.M., and Mutschler, M.A. (1993). Acylsugars of *Lycopersicon pennellii* deter settling and feeding of the green peach aphid (Homoptera, Aphididae). J. Econ. Entomol. **86**: 34-39.

Sawada, Y., Kuwahara, A., Nagano, M., Narisawa, T., Sakata, A., Saito, K., and Hirai, M.Y. (2009). Omics-based approaches to methionine side chain elongation in Arabidopsis: characterization of the genes encoding methylthioalkylmalate isomerase and methylthioalkylmalate dehydrogenase. Plant Cell Physiol. **50**: 1181-1190.

Scheel, D., and Casida, J.E. (1985). Acetohydroxyacid Synthase Inhibitors as Herbicides (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 344-355.

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

Schilmiller, A.L., Gilgallon, K., Ghosh, B., Jones, A.D., and Last, R.L. (2016). Acylsugar acylhydrolases: Carboxylesterase-catalyzed hydrolysis of acylsugars in tomato trichomes. Plant Physiol. **170**: 1331-1344.

Schilmiller, A.L., Moghe, G.D., Fan, P., Ghosh, B., Ning, J., Jones, A.D., and Last, R.L. (2015). Functionally divergent alleles and duplicated loci encoding an acyltransferase contribute to acylsugar metabolite diversity in *Solanum* trichomes. Plant Cell **27**: 1002-1017.

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

Schutt, B.S., Abbadi, A., Loddenkotter, B., Brummel, M., and Spener, F. (2002). Beta-ketoacyl-acyl carrier protein synthase IV: a key enzyme for regulation of medium-chain fatty acid synthesis in *Cuphea lanceolata* seeds. Planta **215**: 847-854.

Senthil-Kumar, M., and Mysore, K.S. (2011). New dimensions for VIGS in plant functional genomics. Trends Plant Sci. **16**: 656-665.

Senthil-Kumar, M., and Mysore, K.S. (2014). Tobacco rattle virus-based virus-induced gene silencing in *Nicotiana benthamiana*. Nat. Protoc. **9**: 1549-1562.

Shapiro, J.A., Steffens, J.C., and Mutschler, M.A. (1994). Acylsugars of the wild tomato *Lycopersicon pennellii* in relation to geographic distribution of the species. Biochem. Syst. Ecol. **22**: 545-561.

Shockey, J.M., Fulda, M.S., and Browse, J. (2003). Arabidopsis contains a large superfamily of acyl-activating enzymes. Phylogenetic and biochemical analysis reveals a new class of acyl-coenzyme a synthetases. Plant Physiol. **132**: 1065-1076.

Singh, V., Roy, S., Giri, M.K., Chaturvedi, R., Chowdhury, Z., Shah, J., and Nandi, A.K. (2013). *Arabidopsis thaliana* FLOWERING LOCUS D is required for systemic acquired resistance. Mol. Plant Microbe Interact. **26**: 1079-1088.

Sinha, N.R., Rowland, S.D., and Ichihashi, Y. (2016). Using gene networks in EvoDevo analyses. Curr. Opin. Plant Biol. **33**: 133-139.

Slabaugh, M.B., Leonard, J.M., and Knapp, S.J. (1998). Condensing enzymes from *Cuphea wrightii* associated with medium chain fatty acid biosynthesis. Plant J. **13**: 611-620.

Slocombe, S.P., Schauvinhold, I., McQuinn, R.P., Besser, K., Welsby, N.A., Harper, A., Aziz, N., Li, Y., Larson, T.R., Giovannoni, J., Dixon, R.A., and Broun, P. (2008). Transcriptomic and reverse genetic analyses of branched-chain fatty acid and acyl sugar production in *Solanum pennellii* and *Nicotiana benthamiana*. Plant Physiol. **148**: 1830-1846.

Sonderby, I.E., Burow, M., Rowe, H.C., Kliebenstein, D.J., and Halkier, B.A. (2010). A complex interplay of three R2R3 MYB transcription factors determines the profile of aliphatic glucosinolates in Arabidopsis. Plant Physiol. **153**: 348-363.

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

Taylor, N.L., Heazlewood, J.L., Day, D.A., and Millar, A.H. (2004). Lipoic acid-dependent oxidative catabolism of alpha-keto acids in mitochondria provides evidence for branched-chain amino acid catabolism in Arabidopsis. Plant Physiol. **134**: 838-848.

Team, R.C. (2014). R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Tholl, D., Chen, F., Petri, J., Gershenzon, J., and Pichersky, E. (2005). Two sesquiterpene synthases are responsible for the complex mixture of sesquiterpenes emitted from Arabidopsis flowers. Plant J. **42**: 757-771.

Thorvaldsdottir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. **14**: 178-192.

Tian, J., Pei, H., Zhang, S., Chen, J., Chen, W., Yang, R., Meng, Y., You, J., Gao, J., and Ma, N. (2014). TRV-GFP: a modified Tobacco rattle virus vector for efficient and visualizable analysis of gene function. J. Exp. Bot. **65**: 311-322.

Torjek, O., Witucka-Wall, H., Meyer, R.C., von Korff, M., Kusterer, B., Rautengarten, C., and Altmann, T. (2006). Segregation distortion in Arabidopsis C24/Col-0 and Col-0/C24 recombinant inbred line populations is due to reduced fertility caused by epistatic interaction of two loci. Theor. Appl. Genet. **113**: 1551-1561.

Torjek, O., Meyer, R.C., Zehnsdorf, M., Teltow, M., Strompen, G., Witucka-Wall, H., Blacha, A., and Altmann, T. (2008). Construction and analysis of 2 reciprocal Arabidopsis introgression line populations. J. Hered. **99**: 396-406.

van der Hoeven, R.S., and Steffens, J.C. (2000). Biosynthesis and elongation of short- and medium-chain-length fatty acids. Plant Physiol. **122**: 275-282.

Vijay, N., Poelstra, J.W., Kunstner, A., and Wolf, J.B. (2013). Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive in silico assessment of RNA-seq experiments. Mol. Ecol. **22**: 620-634.

Walters, D.S., and Steffens, J.C. (1990). Branched chain amino acid metabolism in the biosynthesis of *Lycopersicon pennellii* glucose esters. Plant Physiol. **93**: 1544-1551.

Wentzell, A.M., Rowe, H.C., Hansen, B.G., Ticconi, C., Halkier, B.A., and Kliebenstein, D.J. (2007). Linking Metabolic QTLs with network and cis-eQTLs controlling biosynthetic pathways. PLOS Genetics **3**: e162.

Zerbino, D.R., and Birney, E. (2008). Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. **18**: 821-829.

Zhao, J., Williams, C.C., and Last, R.L. (1998). Induction of Arabidopsis tryptophan pathway enzymes and camalexin by amino acid starvation, oxidative stress, and an abiotic elicitor. Plant Cell **10**: 359-370.

Zhou, X.B., Lindsay, H., and Robinson, M.D. (2014). Robustly detecting differential expression in RNA sequencing data using observation weights. Nucleic Acids Res. **42**.

Zhu, M., Jiang, L., Bai, B., Zhao, W., Chen, X., Li, J., Liu, Y., Chen, Z., Wang, B., Wang, C., Wu, Q., Shen, Q., Dinesh-Kumar, S.P., and Tao, X. (2017). The intracellular immune receptor Sw-5b confers broad-spectrum resistance to tospoviruses through recognition of a conserved 21-amino acid viral effector epitope. Plant Cell **29**: 2214-2232.