GENOMIC APPROACHES TO IDENTIFYING ACYLSUGAR METABOLISM GENES IN *SOLANUM PENNELLII*

A Dissertation

by

WANGMING JI

Submitted to the Graduate and Professional School 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,	Alan Pepper
	Timothy Devarenne
	Wayne Versaw
Head of Department,	Thomas McKnight

August 2021

Major Subject: Biology

ABSTRACT

Acylsugars are specialized metabolites secreted by plants of the nightshade family. These chemicals have documented insecticidal characteristics. To make acylsugars, the plants attach branched-chain and straight-chain fatty acyl groups to sugar backbones. While many steps of acylsugar synthesis pathways have been identified, several acylsugar metabolic genes remain unknown. In this study, we used comparative genomic approaches in two independent experiments: a differential gene expression analysis between high- and low-acylsugar-producing *S. pennellii* accessions, and another differential gene expression analysis between 10 high- and 10 low-acylsugar-producing F2 plants of *S. pennellii* LA 0716 and *S. lycopersicum* cv. VF36 cross, to further investigate the biosynthesis of acylsugar in *S. pennellii*. Differential gene expression analysis tools were used to identify known and novel candidate genes, including genes putatively encoding fatty acid synthases, acyl-activating enzymes, ATP binding cassette (ABC) transporters, and CO_2 fixation proteins, that were positively correlated with acylsugar accumulation. When *S. pennellii* leaves were treated with an inhibitor of BCAA biosynthesis, expression of known and candidate genes was repressed in response to inhibitor in a concentration-dependent manner. We identified four genes, *Sopen05g009610*, *Sopen07g006810*, *Sopen05g032580* and *Sopen05g034770* that were common in two differentially expressed gene (DEG) sets collected from the two independent comparative transcriptomics experiments and were under positive selection. I confirmed two candidate genes, *Sopen05g009610*, encoding a component of a fatty

acid synthase, and *Sopen07g006810*, encoding a small subunit of Rubisco, through virus-induced gene silencing (VIGS), to be involved in medium-chain fatty acids synthesis pathway and carbon fixation associated with acylsugar production, respectively. These results provide strong support for the involvement of novel candidate genes we identified in acylsugar biosynthesis, and validate our approach of combining differential gene expression and evolutionary analysis.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. McKnight, and my committee members, Dr. Devarenne, Dr. Pepper and Dr. Versaw, for their guidance and support throughout the course of this research.

Thanks also go to my lab member, Sachi, and colleagues and the department faculty and staff for making my time at Texas A&M University a wonderful experience.

Finally, thanks to my mother and father and for their encouragement and to my girlfriend for her patience and love.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Dr. Thomas McKnight [advisor] and Dr. Alan Pepper of the Department of the Biology, Dr. Timothy Devarenne of the Department of Biochemistry and Biophysics, and Dr. Wayne Versaw of the Department of the Biology.

The comparative transcriptomics between high- and low-acylsugar-producing accessions of *Solanum pennellii* experiment and transcriptome analysis after biochemical inhibition of BCAA biosynthesis by imazapyr experiment were collaborations with lab member Dr. Sabyasachi Mandal. All other work conducted for the dissertation was completed by the student independently.

Graduate study was supported by a teaching assistantship from Texas A&M University and, in part, from a research grant funded by USDA (Grant No. 2011-38821-30891).

NOMENCLATURE

BCAA	Branched-Chain Amino Acid
BCFA	Branched-Chain Fatty Acid
CCM	Central Carbon Metabolism
DEGs	Differentially Expressed Genes
dN	Nonsynonymous substitution rate
dS	Synonymous substitution rate
dN/dS	Nonsynonymous to synonymous ratio
FAS	Fatty Acid Synthase
FPKM	Fragments Per Kilobase of transcript per Million mapped reads
GC-MS	Gas Chromatography-Mass Spectrometry
LC-MS	Liquid Chromatography–Mass Spectrometry
LDW	Leaf Dry Weight
PAML	Phylogenetic Analysis by Maximum Likelihood
RBH	Reciprocal best BLAST hit
SCFA	Straight-Chain Fatty Acid
TRV	Tobacco Rattle Virus
VIGS	Virus-Induced Gene Silencing
WGCNA	Weighted correlation network analysis

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION AND LITERAURE REVIEW

Plants produce large numbers of metabolites with wide structural and functional diversity, including primary metabolites essential for highly conserved core processes for growth and development, as well as small molecules known as specialized or secondary metabolites, which contribute to plant-environment interactions such as pathogen defense and pollinator attraction (Mithöfer, 2012; Pichersky, 2011; Fiehn, 2002; Afendi et al., 2012). Unlike well understood primary metabolic pathways, biosynthesis pathways of many specialized metabolites remain to be fully elucidated. Enzymes in biosynthetic pathways of specialized metabolites primarily arise through gene duplication, which appear to evolve faster than those producing primary metabolites, resulting in cell- and tissue-specific gene expression and novel enzymatic activity (Pichersky and Lewinsohn, 2011; Moghe, 2015; Panchy et al., 2016).

Acylsugars are nonvolatile and viscous secondary metabolites secreted through glandular trichomes of multiple species in the Solanaceae (Fobes et al., 1985; King et al., 1990; Moghe et al., 2017). These compounds contribute directly and indirectly to plant-defense, such as providing resistance against insect herbivores (Puterka et al., 2003; Alba et al., 2009; Leckie et al., 2016), and by mediating multitrophic defense by attracting predators of herbivores through volatile short-chain aliphatic acids produced by the breakdown of acylsugars (Weinhold and Baldwin, 2011), and protecting plants from microbial pathogens (Shepherd et al., 2005). Acylsugars also function in retaining

water by reducing surface tension of adsorbed dew to increase absorption by the leaves, and reducing water loss by acting as a thickened boundary (Fobes et al., 1985). In addition, acylsugars and related synthetic compounds have been applied as food, cosmetic and pharmaceutical products surfactants (Hill and Rhode, 1999; Dembitsky, 2004), proving their commercial and medicinal value. All these beneficial properties led to interest in understanding acylsugars biosynthesis and factors that control their amount for breeding agronomically important crops with better insect resistance (Mutschler and Wintermantel, 2006).

Solanum pennellii, a relative of the cultivated tomato *Solanum lycopersicum*, is endemic to arid western slopes of the Peruvian Andes (Warnock, 1991), and it has the ability to withstand drought conditions (Yu, 1972; Martin et al., 1988). Acylsugars secreted from type-IV glandular trichomes in *S. pennellii* represent a remarkably large fraction, up to 20%, of the total dry weight of *S. pennellii* leaves (Fobes et al., 1985; Slocombe et al., 2008). In contrast, *S. lycopersicum* produces extremely low levels of acylsugars (Fobes et al., 1985). Acylsugars typically consist of branched and/or straight fatty acyl chains esterified to hydroxyl groups of glucose or sucrose molecules. A large portion of acylsugars from *S. pennellii* is a mixture of 2,3,4-tri-*O*-acylated glucose esters possessing three C4 to C12 hydrocarbon chains, while the rest are acylsucroses, which are composed of a sucrose backbone with three C4 to C12 aliphatic acyl groups esterified only on the pyranose ring (Walters and Steffens, 1990; Shapiro et al., 1994; Schilmiller et al., 2015; Fan et al., 2020; Figure 1A). *S. lycopersicum* only produces tri-

and tetra-*O*-acylsucrose esters with acyl groups ranging from C2 to C12 (Schilmiller et al. 2015).

S. pennellii also has the potential to become a novel biofuel feedstock. The huge environmental impacts that fossil fuels have, including everything from acid rain to global warming, are difficult to avoid (Ndimba et al., 2013). Therefore, the use of sustainable energy has become one of the most pressing concerns of the 21st century. Biofuels, unlike fossil fuels, are renewable energy sources, and they do not contribute to a net increase of atmospheric CO_2. However, biofuel feedstock has to be grown and there is only so much land in the world fit for agriculture, which means growing biofuels feedstock could detract from the process of growing food (Borak et al., 2013). Thus, developing low-input species that can be grown on non-traditional agricultural lands to supply feedstock for biofuel production is an essential priority. With a quick ethanol rinse and concentration, *S. pennellii* acylsugars can be easily collected since they accumulate on the *S. pennellii* leaf surface. This process provides an immediate densification of energy and value. Through a simple transesterification reaction with alcohol and a strong base, the glucolipid yields one molecule of glucose and three molecules of C4 to C12 fatty acid esters. These esters have carbon chains within the range of bio-gasoline and are predicted to be compatible with current fuel technology and gasoline engines.

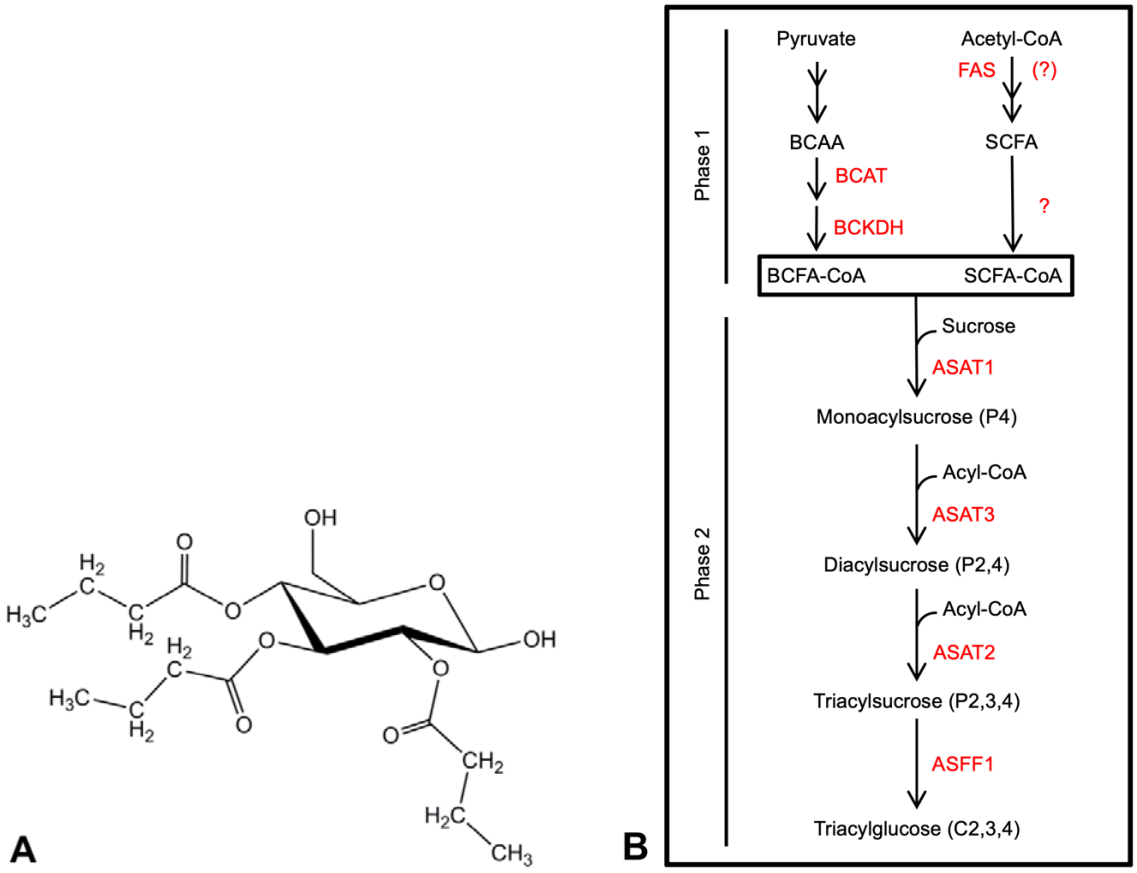

Figure 1. Acylglucose and current model of acylsugar production in *S. pennellii.* (A) Structure of acylglucose in *S. Pennellii*. **(B)** Branched-chain fatty acids (BCFAs) are derived from branched-chain amino acids (BCAAs), and straight-chain fatty acids (SCFAs) are assumed to be produced by fatty acid synthase (FAS). Acylsugar acyltransferases (ASATs) use fatty acyl-CoAs for acylsucrose biosynthesis. Recently, an invertase-like enzyme (acylsucrose β-fructofuranosidase 1; ASFF1) was reported, which produces acylglucose from acylsucrose. Enzymes are highlighted in red. Unidentified enzymes and transporters are marked with '?'. Double arrows indicate more than one enzymatic step. C2-4 indicates esterification at respective positions on glucose. P2-4 indicates esterification at respective positions on the sucrose pyranose ring (Mandal et al., 2020).

Biosynthesis of acylsugars can be broadly divided into two phases: 1) synthesis of fatty acyl chains, 2) esterification of fatty acyl chains to glucose or sucrose backbones (Figure 1B). The fatty acyl substituents on acylsugars are composed of branched and straight-chain fatty acids (Burke et al., 1987; King et al., 1990; Walters and Steffens, 1990). For *S. pennellii* acylsugars, the primary branched-chain fatty acids (BCFAs) side chains are 2-methyl-propanoic acid (C4), 3-methyl-butanoic acid (C5), 2-methyl-butanoic acid (C5), and 8-methyl-nonanoic acid (C10) derived from branched-chain amino acids (BCAAs), such as Val, Ile, and Leu (Kandra et al., 1990; Walters and Steffens, 1990). Branched medium-chain acyl groups are produced from the branched short-chain precursors through elongation by units of two carbons via a special fatty acid synthase mechanism in the case of *S. pennellii* LA 0716 (van der Hoeven and Steffens, 2000; Kroumova and Wagner, 2003). Primary straight-chain fatty acids (SCFAs) side chains are n-decanoic acid (C10) and n-dodecanoic acid (C12) which arise from acetate and are extended to the final length through a de novo fatty acid biosynthetic process with a chain elongation process of two-carbon extensions mediated by the fatty acid synthase (FAS) complex (Walters and Steffens, 1990).

A previously proposed acylglucose biosynthetic pathway in *S. pennellii* invoked that the attachment of acyl chains to the glucose backbone starts with the conjugation of fatty acids and UDP-glucose catalyzed by UDP-glucose:fatty acid glucosyltransferase (UDP-Glc:FA GT) to synthesize a high-energy intermediate 1-*O*-acyl-β-D- glucose, which serves as an acyl donor in the subsequent acyltransfer reactions (Ghangas and Steffens, 1993; Kuai et al., 1997; Li et al., 1999). Then, two equivalents of activated 1-

O-acyl-β-D- glucose form a 1,2-substituted diacyl-glucose and a glucose through disproportionation under the catalysis of a serine carboxypeptidase-like glucose acyltransferase (SCPL GAT) (Li et al., 1999; Li and Steffens, 2000). Since this acyltransferase does not appear to produce tri- or tetra-acylated derivatives in vitro (Li et al., 1999), synthesis of the final product, 2,3,4-tri-O-acylglucose requires removal of acyl chains from C1 position, and esterification/transfer of acyl chains to C3 and C4 positions of glucose by uncharacterized acyl-transferase utilizing 1-*O*-acyl-β-D-glucose or fatty acyl CoA as acyl donors. However, there is no genetic evidence to support this model in acylsugar biosynthesis, and my studies do not support a role for these enzymes in acylsugar production.

Acylglucoses are produced from acylsucroses in *S. pennellii* (Schilmiller et al., 2015; Fan et al., 2016; Fan et al., 2017; Leong et al., 2019; Fan et al., 2020). In *S. lycopersicum,* fatty acyl chains are added to both pyranose and furanose rings of sucrose by four BADH family acylsugar acyltransferases (ASATs) to produce tri- and tetra-acylated sucrose (Schilmiller et al., 2012; Schilmiller et al., 2015; Fan et al., 2016; Fan et al., 2017). In *S. pennellii*, three ASATs add acyl chains only to pyranose ring of sucrose backbone to yield pyranose acylated tri-acylsucroses, most of which are subsequently hydrolyzed into tri-acylglucose by an invertase-like enzyme, acylsucrose β-fructofuranosidase (ASFF1) (Schilmiller et al., 2015; Fan et al., 2016; Fan et al., 2017; Leong et al., 2019).

Although several acylsugar metabolic genes have been identified, the regulation of acylsugar synthesis pathway has not been well characterized, especially with regard to

how *S. pennellii* accumulates such a high level of acylsugars compared to *S. lycopersicum*. The levels of total acylsugars vary significantly among different accessions of *S. pennellii* (Shapiro et al., 1994). *S. pennellii* accessions LA0716, LA1941 and LA1946 produce high amounts of acylsugar, LA1302 produces medium amount of acylsugar, while *S. pennellii* accessions LA1911, LA1912, LA1920 and LA1926 lack detectable levels of acylsugar production, less than 1% of leaf dry weight (Shapiro et al., 1994). In this study, we conducted transcriptomic comparisons between high- and low-glucolipid producing accessions of *S. pennellii*, as well as transcriptomic comparisons between high- and low-acylsugar producing F2 population from a *S. pennellii* LA 0716 and *S. lycopersicum* cv. VF36 cross to identify candidate biosynthetic and regulatory genes related to acylsugar metabolism. Comparative transcriptomics after biochemical inhibition of BCAA biosynthesis, genome-scale dN/dS ratio estimation of *S. pennellii* and *S. lycopersicum* putative orthologs and virus-induced gene silencing (VIGS) further refined the list of candidate genes and illuminated possible regulatory mechanisms of acylsugar metabolic pathway. This analysis revealed two genes, *SpKAR*, encoding a FAS component, and *SpRbcS*, encoding a small subunit of Rubisco, that showed differential expression in high acylsugar producing line, were preferentially expressed in trichomes relative to the underlying tissue, and showed rapid evolution. These results extend our understanding of acylsugar biosynthesis and regulatory mechanisms required by the breeding strategies for engineering acylsugar quantities in crops.

Specific Aim 1: Identification of differentially expressed genes between high- and low-acylsugar-producing *S. pennellii* accessions via comparative transcriptomics

Methods: RNA-seq methods allowed us to collect transcriptome profiles from different *S. pennellii* accessions. Comparative transcriptomics between high- and low-acylsugar-producing accessions of *S. pennellii* identified differentially expressed genes between these two groups. Transcriptome analysis after biochemical inhibition of BCAA biosynthesis by imazapyr further implied possible regulatory mechanisms of acylsugar production.

Approach: mRNA profiles were collected for high- and low-acylsugar-producing accessions of *S. pennellii*. Differential gene expression analysis tools were used to identify known and novel candidate genes, including genes putatively encoding fatty acid synthases, acyl-activating enzymes, ATP binding cassette (ABC) transporters, and CO_2 fixation proteins, that were positively correlated with acylsugar accumulation, except two genes previously reported to be involved in acylglucose biosynthesis. When *S. pennellii* leaves were treated with an inhibitor of BCAA biosynthesis, expression of know and candidate genes was repressed in response to inhibitor in a concentration-dependent manner.

Specific Aim 2: Transcriptomic comparation between "HIGH"- and "LOW"- F2 accessions combined with genome-scale dN/dS ratio estimation of *S. pennellii* and *S. lycopersicum* putative orthologs to refine the list of candidate genes involved in acylsugar metabolism

Methods: We observed that there was a segregation on the acylsugar production amount phenotype for the F2 population between *S. pennellii* LA 0716 and *S. lycopersicum* cv. VF36. RNA-seq was performed on 10 highest- and 10 lowest-acylsugar-producing F2, and differentially expressed genes were identified. Putative *S. lycopersicum* orthologs of *S. pennellii* genes were uncovered using reciprocal BLAST between all annotated *S. pennellii* and *S. lycopersicum* genes. Genome-scale dN/dS ratios were measured on putative *S. pennellii* and *S. lycopersicum* orthologs pairs.

Approach: To identify genes that were common in two DEG sets collected from the two independent comparative transcriptomics experiments and were under positive selection, a Venn diagram was created to calculate the intersections of these three gene sets. The output indicated that *Sopen05g009610*, *Sopen07g006810*, *Sopen05g032580* and *Sopen05g034770* exist in all three sets.

Specific Aim 3: *In vivo* functional validation of selected candidate genes

Methods: We used virus-induced gene silencing method to functional validate two candidate genes preferentially expressed in trichomes, *Sopen05g009610* (*SpKAR*) and *Sopen07g006810* (*SpRbcS*), which were hypothesized to be involved in acylsugar related medium chain fatty acid synthesis pathway and carbon fixation, respectively.

Phylogenetic analyses were performed to uncover distinct clades of SpKAR and SpRbcS sequences in solanaceous species.

Approach: Quantitative reverse transcription PCR (qRT-PCR) results confirmed that *SpKAR* and *SpRbcS* are highly expressed in *S. pennellii* LA0716 trichomes comparing with underlying shaved stem tissue. These two genes were targeted in *S. pennellii* LA0716 for VIGS using tobacco rattle virus (TRV)-based silencing vectors. Extracts of trichome and leaf surface metabolites from these VIGS lines were analyzed by liquid chromatography–mass spectrometry (LC-MS) and gas chromatography–mass spectrometry (GC-MS). Transcript enrichment in trichomes of *SpKAR* and *SpRbcS*, together with VIGS assay validation, indicated they were indeed involved in acylsugar biosynthesis. Phylogenetic positions of SpKAR and SpRbcS indicated distinct secondary metabolism subclades reside outside of the primary metabolism subclades, suggesting duplication events leading to different clades of the phylogenetic tree related to evolution of acylsugar biosynthesis activities.

Impact

Increasing acylsugar-mediated insect resistance has been a target of various breeding programs in solanaceous crops. Unraveling the complex network of acylsugar biosynthesis and factors controlling acylsugar amount would be essential for breeding agronomically important crops with better insect resistance. In this study, we exploited variation in acylsugar production among different *S. pennellii* accessions and within a segregating F2 population from *S. pennellii* LA 0716 and *S. lycopersicum* cv. VF36

cross, through a combination of comparative transcriptomics and evolutionary analysis to identify a number of genes with known and putative functions in acylsugar metabolism. Additionally, we used virus-induced gene silencing methods to confirm the proposed roles of two candidate genes in acylsugar metabolism. Overall, data presented here should serve as a valuable resource in future studies of acylsugar metabolism in *S. pennellii* and other plants.

Perspectives:

In all, this dissertation contains three major experiments:

1) Comparative transcriptomics study between high- and low-acylsugar-producing accessions of *S. pennellii*.

2) Comparative transcriptomics study between low- and high-acylsugar-producing F2 population of *S. pennellii* LA 0716 and *S. lycopersicum* cv. VF36 and genome-wide dN/dS ratio estimation of *S. pennellii* and *S. lycopersicum* putative orthologs.

3) Functional validation of two candidate genes involved in acylsugar biosynthesis.

The first experiment was complete and published in Plant Cell in January, 2020. The second and third experiment refined the list of candidate genes potentially involved in acylsugar biosynthesis, and confirm that two candidate genes were indeed involved in acylsugar metabolism. The aim of this dissertation is to extend our understanding of the genetic and biochemical mechanisms leading to acylsugar production in *S. pennellii*, which may contribute to studying how plants evolved to mediate environmental interactions and developing pests resistant tomato breeding lines.

CHAPTER II

RESULTS

Expression profiles comparison between high- and low-acylsugar-producing *S. pennellii* accessions

A differential gene expression analysis was performed between high-acylsugar-producing accessions group (LA1941, LA1946, and LA0716; referred to as "HIGH" group) and low-acylsugar-producing accessions group (LA1911, LA1912, and LA1926; referred to as "LOW" group) of *S. pennellii* (Table 1) (Shapiro et al., 1994). We collected leaf samples from three individual plants of each accession, as biological replicates. An average of more than 30 million 125-bp paired-end reads for each sample were used for the comparative transcriptomics analysis. A total of 19,379 genes passed filtering criteria for minimum expression levels (see Methods), and 1679 differentially expressed genes (DEGs) were obtained with edgeR (Supplemental Dataset 1: Sheet 1). Within the 1679 DEGs, 931 genes were upregulated and 748 genes were downregulated in the 'HIGH' group.

We also applied an independent differential gene expression analysis between the high-acylsugar-producing accession LA0716 and another low-acylsugar-producing accession LA1920, with four biological replicates in each accession and similar sequencing coverage. A total of 19,967 genes passed filtering criteria for minimum expression levels (see Methods) in this analysis, and 3524 DEGs were identified by edgeR package, with 1465 upregulated and 2059 downregulated genes in LA0716

(Supplemental Dataset 1: Sheet 2). Accession-specific DEGs from this differentially expression analysis may explain the higher number of DEGs identified by LA0716 vs. LA1920 comparison than the DEG numbers identified by "HIGH" vs. "LOW" comparison. We compared "HIGH" vs. "LOW" DEG set with LA0716 vs. LA1920 DEG set and obtained 1,087 DEGs that were common to both comparisons, with 586 genes upregulated and 501 genes downregulated in high-acylsugar-producing accessions (Supplemental Dataset 1: Sheet 3).

Table 1. Amount of total acylsugars produced by different accessions of *Solanum pennellii*. Amounts are represented as mean percentage of leaf dry weight (% LDW) from 12-week-old plants (n=8).

Accession	Designation	% LDW	Std. Dev.
LA1911	LOW	0.45	0.09
LA1912	LOW	0.69	0.13
LA1920	LOW	0.55	0.23
LA1926	LOW	0.65	0.17
LA1302	MEDIUM	5.62	0.96
LA1941	HIGH	12.21	0.63
LA1946	HIGH	12.57	0.35
LA0716	HIGH	13.92	0.43

Furthermore, we obtained the transcriptome profile of an intermediate-acylsugar-producing accession LA1302 (referred to as "MEDIUM" group) with leaf samples from three biological replicates (Table 1) (Shapiro et al., 1994). By comparing the acylsugar

candidate gene expression levels in “HIGH”, “MEDIUM” and “LOW” group, it is possible to estimate the relation between candidate genes expression levels and the amount of acylsugar produced. Results for selected genes are shown in Table 2.

Table 2. Differentially expressed genes involved in acylsugar production. 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 (Mandal et al., 2020).

Gene ID	“HIGH” vs. “LOW”		LA1920 vs. LA0716		Annotation
	Log_2 FC	FDR	Log_2 FC	FDR	
Branched-chain amino acid (BCAA)/ Branched-chain fatty acid (BCFA) metabolism					
Sopen11g004560	1.44	4.30E-05	1.63	2.20E-05	Acetolactate synthase small subunit
Sopen07g027240	1.46	1.00E-11	1.38	2.80E-06	Ketol-acid reductoisomerase
Sopen05g032060	1.78	1.80E-10	1.71	4.50E-06	Dihydroxy-acid dehydratase
Sopen08g005060	1.01	3.10E-02	1.38	3.60E-07	Isopropylmalate synthase
Sopen08g005140	4.97	3.30E-13	5.12	6.40E-14	Isopropylmalate synthase
Sopen04g030820	5.28	3.50E-13	4.55	4.40E-09	Branched-chain aminotransferase-2
Sopen04g026270	3.37	2.80E-12	2.66	2.10E-07	Branched-chain keto acid dehydrogenase E1 subunit
Sopen01g028100	1.99	1.40E-16	1.53	9.80E-08	Branched-chain keto acid dehydrogenase E2 subunit
Sopen07g023250	4.75	1.40E-08	3.87	2.30E-04	3-Hydroxyisobutyryl-CoA hydrolase
Sopen05g023470	1.34	1.40E-02	1.08	1.00E-02	3-Hydroxyisobutyryl-CoA hydrolase
Sopen12g032690	2.02	7.80E-03	1.87	1.30E-02	Mitochondrial acyl-CoA thioesterase

Gene ID	"HIGH" vs. "LOW"		LA1920 vs. LA0716		Annotation
	Log_2 FC	FDR	Log_2 FC	FDR	
Fatty acid synthase (FAS) components; * misannotated as two separate genes					
Sopen12g004230 *	5.28	4.40E-18	5.19	1.40E-12	Beta-ketoacyl-ACP synthase II (N-terminus)
Sopen12g004240 *	4.73	1.50E-26	4.39	8.90E-14	Reverse transcriptase; KAS IV/ KAS II-like domain (C-terminus)
Sopen08g002520	4.56	1.40E-13	3.67	3.70E-07	Beta-ketoacyl-ACP synthase III
Sopen05g009610	3.32	3.80E-11	2.73	1.80E-08	Beta-ketoacyl-ACP reductase
Sopen12g029240	3.34	1.60E-09	2.85	2.90E-08	Enoyl-ACP reductase domain
Acyl-activating enzymes					
Sopen02g027670	3.08	5.50E-10	2.49	2.70E-05	Acyl-activating enzyme 1
Sopen02g027680	5.29	5.90E-12	4.23	3.70E-06	Acyl-activating enzyme 1
Sopen07g023200	4.16	2.80E-07	3.72	5.30E-05	Acyl-activating enzyme 1
Sopen07g023220	5.81	3.80E-27	4.02	3.60E-06	Acyl-activating enzyme 1

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

Gene ID	"HIGH" vs. "LOW"		LA1920 vs. LA0716		Annotation
	Log_2 FC	FDR	Log_2 FC	FDR	
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

Transporters					
Sopen04g023150	-1.16	2.00E-02	-1.82	4.40E-09	ABC transporter F family member
Sopen01g047950	-1.68	2.10E-02	-2.82	3.30E-04	ABC transporter G family member
Sopen04g005380	3.19	2.80E-08	2.8	6.50E-04	ABC transporter G family member
Sopen12g034820	4.37	2.10E-12	3.47	4.00E-06	Pleiotropic drug resistance protein 1-like (ABC transporter G family)
Sopen03g001870	4.52	1.10E-32	4	3.60E-09	ABC transporter B family member
Central carbon metabolism					
Sopen07g006810	5.01	6.70E-12	3.63	4.30E-04	RUBISCO small subunit
Sopen08g020180	2.11	1.90E-10	2.18	2.30E-07	NADP-dependent malic enzyme
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
Transcription factors					
Sopen05g008450	4.9	1.3E-21	4.15	6.3E-10	AP2 domain

Gene ID	"HIGH" vs. "LOW"		LA1920 vs. LA0716		Annotation
	Log_2 FC	FDR	Log_2 FC	FDR	
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

Identifying candidate genes involved in acylsugar metabolism from DEG set

Branched and straight fatty acyl chains are first synthesized and then acylated on the pyranose R2, R3, and R4 positions of sucrose (Leong et al., 2019; Fan et al., 2019). P-type acylsucroses are then cleaved to produce acylglucoses (Leong et al., 2019). Because acylsugar BCFAs derive from BCAA (Walters and Steffens, 1990), we first searched for genes involved in BCAA metabolism in the 1,087 DEG set. We found many genes involved in biosynthesis of BCAAs, such as genes encoding acetolactate synthease (ASL), ketol-acid reductoisomerase (KARI) and dihydroxyacid dehydratase (DHAD) were upregulated in high-acylsugar-producing accessions (Table 2). Genes putatively encoding branched-chain aminotransferase (BCAT) and branched-chain keto acid dehydrogenase complex (BCKDH) that convert BCAAs to BCFA-CoAs were also

upregulated in high-acylsugar-producing accessions comparing to low-acylsugar-producing accessions (Table 2; Figure 2). Most of the DEGs related to BCAA and BCFA metabolism showed intermediate levels of expression in the "MEDIUM" accession LA1302, which was consistent with the acylsugar production amount (Figure 3A).

SCFAs arise from acetate and extended to the final length through a de novo fatty acid biosynthetic process mediated by the fatty acid synthase (FAS) complex (Walters and Steffens, 1990). Fatty acids synthesis genes, such as 3-ketoacyl-ACP synthase (KAS), 3-ketoacyl-ACP reductase (KAR) and Enoyl-ACP reductase (EAR), were also upregulated in "HIGH" group (Figure 3A, Table 2).

Acyl-activating enzymes (AAEs), which are ATP/AMP-binding proteins that activate different carboxylic acids by forming fatty acyl-CoA molecules from free fatty acids, ATP and CoA, were reported to convert free medium chain fatty acids produced by de novo fatty acid biosynthesis to medium chain acyl-CoAs in acylsugar metabolism (Shockey et al., 2003; Fan et al., 2020). We found four DEGs (*Sopen02g027670*, *Sopen02g027680*, *Sopen07g023200*, and *Sopen07g023220*) putatively encoding acyl-activating enzyme 1 (AAE1), were more highly expressed in high-acylsugar-producing accessions (Figure 3A, Table 2).

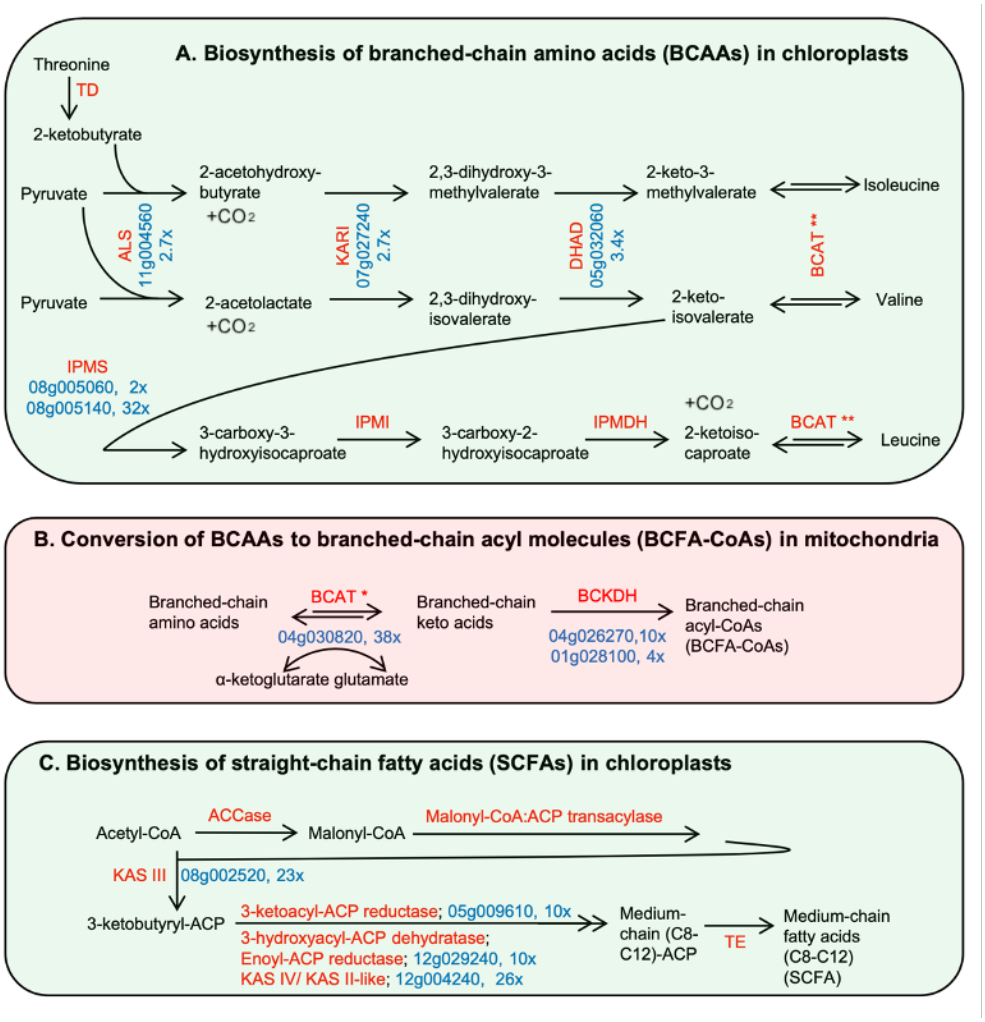

Figure 2. **Biosynthesis of branched-chain (A and B) and straight-chain (C) acyl molecules.** Enzymes are highlighted in red. Gene identifiers of differentially expressed genes (DEGs) and their fold-changes are highlighted in blue. 10X indicates 10-fold higher expression in the 'HIGH' group relative to the 'LOW' group. Conversion of BCAAs to branched-chain acyl molecules is initiated by a transamination reaction, which is reversibly catalyzed by branched- chain aminotransferase (BCAT; marked with an asterisk) enzymes in mitochondria. On the other hand, chloroplast BCATs (marked with double asterisk) catalyze the last step in the biosynthesis of all three BCAAs (Binder, 2010).

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; ACCase= acetyl-CoA carboxylase; KAS= beta- ketoacyl-ACP synthase; TE= thioesterase; ACP= acyl carrier protein (Mandal et al., 2020).

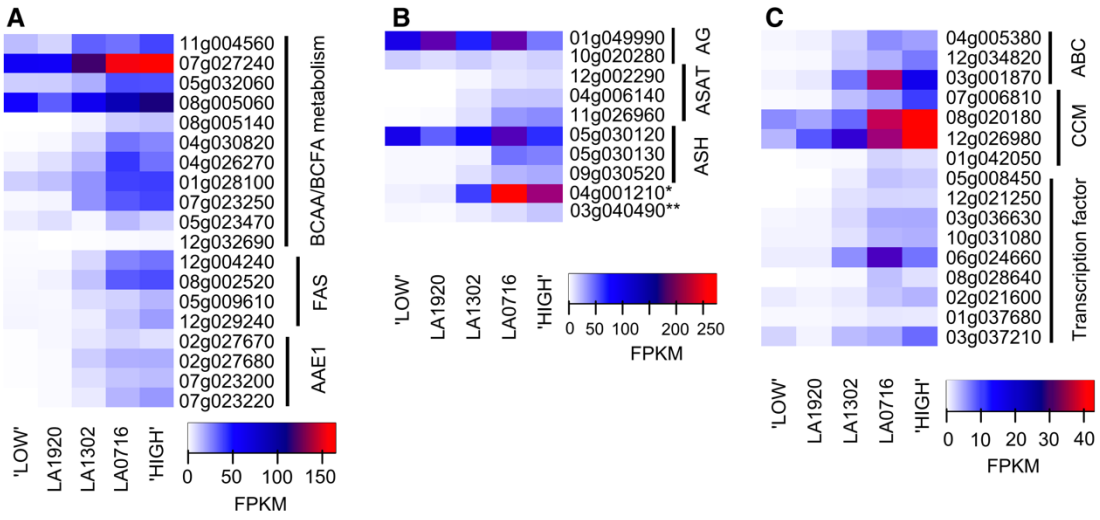

Figure 3. Heatmaps showing expression levels of genes with known and putative functions in acylsugar metabolism. Genes are designated by their gene identifier numbers (Sopen IDs). **(A)** Acylsugar phase 1-related genes. **(B)** Acylsugar phase 2-related genes. *Sopen04g001210* (marked with an asterisk) was identified as a carboxylesterase gene related to ASH (Schilmiller et al., 2016). *Sopen03g040490* (marked with double-asterisk) has been recently reported as a trichome-expressed invertase gene that is capable of producing acylglucose from acylsucrose (Leong et al., 2019). **(C)** Other genes related to acylsugar metabolism.

Abbreviations: FPKM= fragments per kilobase of transcript per million mapped reads; BCAA= branched-chain amino acid; FAS= fatty acid synthase components; AAE1= acyl-activating enzyme 1; AG= acylglucose previous model; ASAT= acylsucrose acyltransferase; ASH= acylsugar hydrolase; ABC= ABC transporters; CCM= central carbon metabolism (Mandal et al., 2020).

Three genes encoding acylsucrose acyltransferase (ASATs) which utilize sucrose and acyl-CoA substrates to produce acylsucroses (Schilmiller et al., 2012; Schilmiller et al., 2015; Fan et al., 2015; Fan et al., 2020; Fan et al., 2017), and recently identified invertase (SpASFF1) which cleaves acylsucroses to produce acylglucoses in *S. pennellii* (Leong et al., 2019), were also upregulated in high-acylsugar-producing accessions (Figure 3B, Table 2).

Three genes encoding acylsugar hydrolases (ASHs), which are carboxylesterases that remove acyl groups from acylsucrose molecules in vitro, and a related carboxylesterase gene (*Sopen04g001210*) were reported to result in reduced acylsugar phenotype in *S. lycopersicum* M82 X *S. pennellii* LA0716 introgression lines (Schilmiller et al., 2016). These genes were found in 1,087 DEG set except *Sopen05g030120* (Sp-ASH1). Schilmiller et al., (2016) reported the expression profile of ASH1 in different tissues of *S. lycopersicum* and propose additional non-trichome-localized functions for Sl-ASH1. Our observation that *Sopen05g030120* shows no differential expression between the "HIGH" and "LOW" groups further supports the hypothesis that *Sopen05g030120* may not have a critical role in acylsucrose metabolism. Genes encoding FAS, AAEs, ASATs, ASHs and putative carboxylesterases also showed intermediate levels of expression in the "MEDIUM" accession LA1302 (Figure 3).

Current evidence shows that acylsugar metabolism occurs majorly in trichome tip cells (Fan et al., 2020). However, it remains unclear how acylsugar molecules are transported across the plasma membrane of trichome tip cells to be exported, either by membrane-associated transporter proteins or via vesicular transport. We identified five putative ATP-binding cassette (ABC) transporter family associated genes in the 1,087 DEG set, and three of which, *Sopen12g034820*, *Sopen04g005380*, and *Sopen03g001870*, were upregulated in high-acylsugar-producing accessions (Table 2). *Sopen12g034820* and *Sopen04g005380* belong to the G subfamily of ABC transporters, which were reported to be involved in the deposition of insect-deterring specialized metabolites

across plasma membrane (Kang et al., 2011). *Sopen03g001870* was predicted to encode an ABC-B subfamily transporter.

In high-acylsugar-producing accessions, glandular trichomes synthesize and secrete high volume of acylsugar which represents an astounding metabolic investment (Fobes et al., 1985). Because 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 high- and low-acylsugar-producing accessions. Photosynthetic activities have been observed in tomato and tobacco trichomes secretory cells (Kandra and Wagner, 1988; Pike and Howells, 2002; Balcke et al., 2017), though the levels of photosynthetic activities in trichomes were significantly lower than in the rest of whole leaf (Laterre et al., 2017; Balcke et al., 2017). *Sopen07g006810,* a gene putatively encoding a RUBISCO small subunit was found upregulated in high-acylsugar-producing accessions. Nevertheless, other RUBISCO small subunit genes were not differentially expressed in high-acylsugar-producing accessions, compared to low-acylsugar-producing accessions. *Sopen08g014550,* a putative starch synthase gene was upregulated in the high-acylsugar-producing accessions, possibly to provide carbon for acylsugar biosynthesis, which requires large amounts of photosynthate essentially importing from other leaf tissues due to the insufficient photosynthesis in trichomes (Balcke et al., 2017). One gene (*Sopen08g020180*) putatively encoding NADP-malic enzyme which catalyzes the oxidative decarboxylation of malate to generate pyruvate, CO_2 and NADPH in chloroplast, was upregulated in high-acylsugar-producing accessions (Figure 3C, Table

2). The products of this reaction can be used for BCAA biosynthesis (pyruvate), or SCFA biosynthesis and carbon fixation (CO_2 and NADPH), which may contribute to acylsugar accumulation. The differentially expressed RUBISCO small subunit gene, starch synthase gene, and NADP-malic gene also had intermediate levels of expression in the "MEDIUM" accession LA1302 (Figure 3C).

Expression levels of two previously reported acylglucose biosynthetic genes are not consistent with acylsugar accumulation levels

Acylglucose metabolic pathway was previously described to begin with a glucosyltransferase (*Sopen01g049990*) adding free fatty acids to uridine diphosphate (UDP)–glucose to form 1-*O*-acyl-d-glucose (Ghangas and Steffens, 1993; Kuai et al., 1997). The second step was a disproportionation of two 1-*O*-isobutyryl-D-glucose catalyzed by a serine carboxypeptidase-like (SCPL) acyltransferase (*Sopen10g020280*) yielding one 1,2-*O*-di-isobutyryl-D-glucose (Li et al., 1999; Li and Steffens, 2000). However, two genes in the proposed model above were not upregulated in high-acylsugar-producing accessions (Figure 3B, Table 2).

Trichome-enriched expression of acylsugar candidate genes

Most genes reported to be involved in tomato acylsugar metabolism are preferentially expressed in trichomes relative to the underlying tissue (Schilmiller et al., 2012; Fan et al., 2015; Fan et al., 2020, Mandal et al., 2020). Because acylsugars are secreted by trichome tip cells, it is reasonable to expect that genes involved in acylsugar secretion

will exhibit trichome-enriched expression. We used quantitative reverse transcription PCR (qRT-PCR) to investigate expression levels of selected candidate genes in isolated stem trichomes versus shaved stems, and found that they showed trichome-enriched expression, with 60-to 250-fold for FAS components, 26- to 562-fold for AAE1 genes, 143- to 470-fold for ABC transporter genes, and 16- to 287-fold for central carbon metabolic genes ($P < 0.001$, Welch two-sample t test, Figure 4).

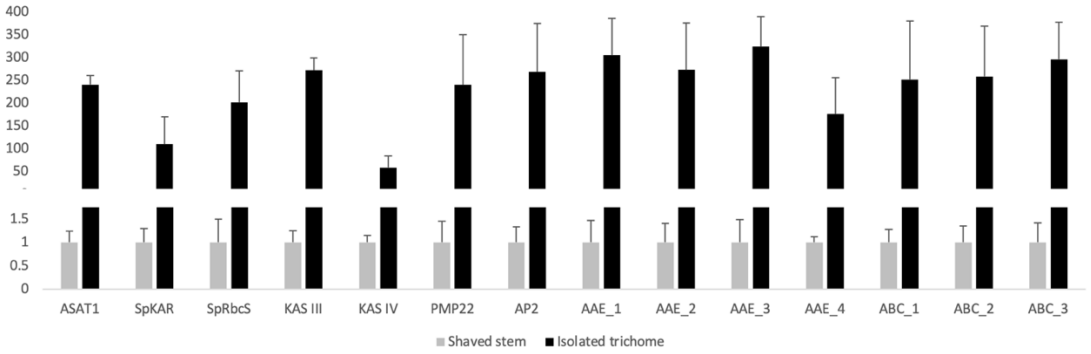

Figure 4. qRT-PCR analysis of the expression levels of seleted genes in trichome. Transcript levels of ASAT1 (*Sopen12g002290*), SpKAR (*Sopen05g009610*), SpRbcS (*Sopen07g006810*), KAS III (*Sopen08g002520*), KAS IV (*Sopen12g004240*), PMP22 (*Sopen11g007710*), AP2 (*Sopen05g008450*), AAE_1 (*Sopen02g027670*), AAE_2 (*Sopen02g027680*), AAE_3 (*Sopen07g023200*), AAE_4 (*Sopen07g023220*), ABC_1 (*Sopen04g005380*), ABC_2 (*Sopen12g034820*), ABC_3 (*Sopen03g001870*) in isolated stem trichomes and shaved stems were measured by qRT-PCR. Error bars indicate standard error (n = 5 individual plants).

Biochemical inhibition of BCAA biosynthesis changes transcript levels of many acylsugar candidate DEGs

The first common step in the biosynthesis of BCAAs is catalyzed by acetolactate synthase (ALS). A previous study showed that biochemical inhibition of this enzyme significantly lowers acylsugar production in *S. pennellii* (Walters and Steffens, 1990). To determine the effect of inhibition of ALS on expression levels of candidate genes

involved in acylsugar metabolism, we treated leaves of the high-acylsugar-producing *S. pennellii* LA0716 with the ALS specific inhibitor imazapyr at 0.1 mM and 1 mM concentrations. We observed many genes involved in BCAA/BCFA metabolism in the 1,087 DEG set showed differential expression patterns in response to imazapyr treatment (Supplemental Dataset 2: Sheet 2). Interestingly, expression levels of branched-chain keto acid dehydrogenase complex E1 and E2 subunits, *Sopen04g026270* and *Sopen01g028100* respectively, were significantly increased under 1 mM imazapyr treatment, possibly due to the increasing amount of 2-ketobutyrate, which is a substrate of ALS that also can be used directly by the dehydrogenase complex to make propionyl-CoA (Walters and Steffens, 1990).

We also observed that FAS components and AAE1 DEGs putatively involved in SCFA metabolism showed significant downregulation at both 0.1 mM and 1 mM imazypyr with a concentration-dependent manner. It has been reported that AAE1 proteins have a peroxisomal location in *Arabidopsis thaliana* (Reumann et al., 2009). *Sopen11g007710*, a gene predicted to encode a PMP22/ Mpv17 family peroxisomal membrane protein, was 7.7-fold and 165-fold downregulated under 0.1 mM and 1 mM imazapyr treatment, respectively (Supplemental Dataset 2: Sheet 2). *Sopen11g007710* was upregulated 29-fold in high-acylsugar-producing accessions (Table 2), and 230-fold higher expression in isolated trichomes than shaved stems (Figure 4), which was consistent with trichome-enriched expression of AAE1 genes (Figure 4). PMP22 is presumably involved in controlling permeability of peroxisomal membrane (Brosius et al., 2002), and these results suggest a role of peroxisome in acylsugar metabolism.

ASATs and the invertase (SpASFF1) genes also showed significant decreases in expression level with a concentration-dependent manner in response to imazapyr treatment, as did ASHs and the related carboxylesterase *Sopen04g001210* (Supplemental Dataset 2: Sheet 2).

Expression levels of three predicted ABC transporters DEGs, as well as DEGs putatively involved in central carbon metabolism, were significantly repressed by imazapyr treatment (Supplemental Dataset 2: Sheet 2). A previous report showed that imazapyr treatment could result in significant induction of nine ABC transporters genes in *Arabidopsis thaliana*, indicating their role in detoxification process (Manabe et al., 2007). We also found 15 putative ABC transporter genes in *S. pennellii* that were upregulated in response to imazapyr either at 1mM concentration or at both 0.1mM and 1mM concentrations (Supplemental Dataset 2: Sheet 2). However, three predicted ABC transporter DEGs were downregulated under imazapyr treatment, consistent with expression profiles of acylsugar metabolic genes.

Sopen01g049990 and *Sopen10g020280*, which encode UDP-Glc:FA GT and SCPL GAT, respectively in previously reported model of acylglucose biosynthesis, showed slightly higher and similar expression levels, respectively in response to imazapyr treatment, consistent with a lack of involvement in acylsugar biosynthesis.

F2 hybrid population between *S. pennellii* LA 0716 and *S. lycopersicum* cv. VF36

To further identify genes involved in acylsugar metabolism, we created a segregating F2 population from an F1 hybrid *S. pennellii* LA 0716 and *S. lycopersicum* cv. VF36. *S.*

pennellii LA 0716 produces large amounts of acylsucroses and acylglucoses, estimated to be as high as 20% of leaf dry weight, while *S. lycopersicum* cv. VF36 produces acylsucroses only, representing up to ~1% of leaf dry weight (Fobes et al., 1985). The F1 progeny of this cross produce ~1% of leaf dry weight acylsugar, close to the *S. lycopersicum* cv. VF36 parent. F2 progeny were generated, and the accumulation of total acylsugar was analyzed across all 114 F2 individuals. Twenty-four F2 plants produce more than 10% acylsugars of leaf dry weight, while fifty-two F2 plants accumulated less than 3% acylsugars of leaf dry weight, indicating that more than one recessive trait is involved in *S. pennellii* acylsugar biosynthesis (Figure 5, Supplemental Dataset 3).

Transcriptomic comparation between "HIGH"- and "LOW"-F2 accessions

We selected the 10 highest-acylsugar-producing F2 individuals with acylsugar amount between 26.90%-14.30% and designated them as the "HIGH"-F2 group. We selected 10 lowest-acylsugar-producing F2 individuals with acylsugar amount between 0.25%-0.84% and designated them as the "LOW"-F2 group (Supplemental Dataset 3). RNA-seq was performed for leaf sample of each plant in "HIGH"-F2 and "LOW"-F2 group, followed by a differential gene expression analysis between "HIGH"- and "LOW"-F2 group (Figure 6). Overall, A total of 20,160 genes were collected after expression filtering (Supplemental Dataset 4: Sheet 1), and 331 differentially expressed genes (DEGs; False Discovery Rate adjusted $P<0.05$, fold change >2; Supplemental Dataset 4: Sheet 2) were identified across all twenty samples. 134 of 331 DEGs were upregulated

and 197 were downregulated in the “HIGH”-F2 group compared with the “LOW”-F2 group.

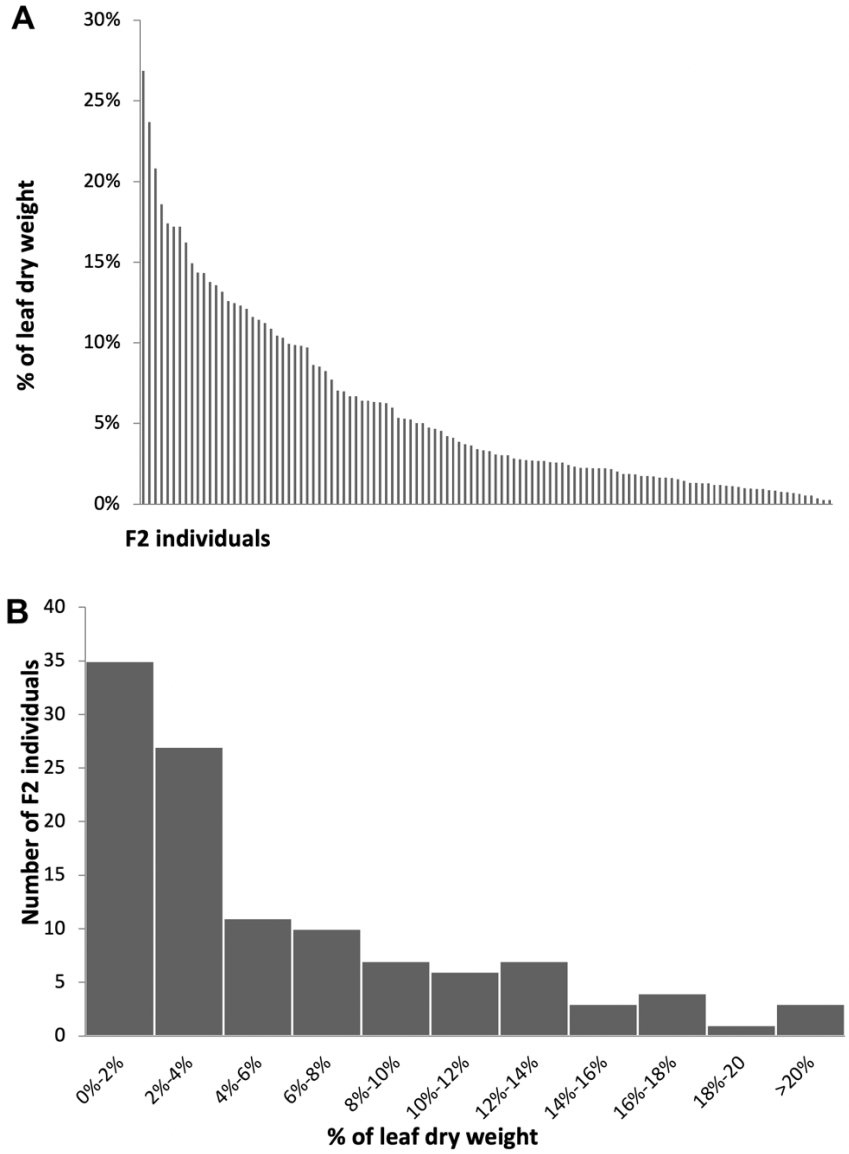

Figure 5. Acylsugar accumulation in *S. pennellii* LA0716 X *S. lycopersicum* VF36 F2 population. (A) Each column represents acylsugar abundance of one *S. pennellii* LA0716 X *S. lycopersicum* VF36 F2 individual. Acylsugar abundance was calculated as a proportion of leaf dry weight from the avenge of three replicates. **(B)** Histogram of acylsugar abundance distribution among *S. pennellii* LA0716 X *S. lycopersicum* VF36 F2 population. Acylsugar abundance was normalized by leaf dry weight from the avenge of three replicates.

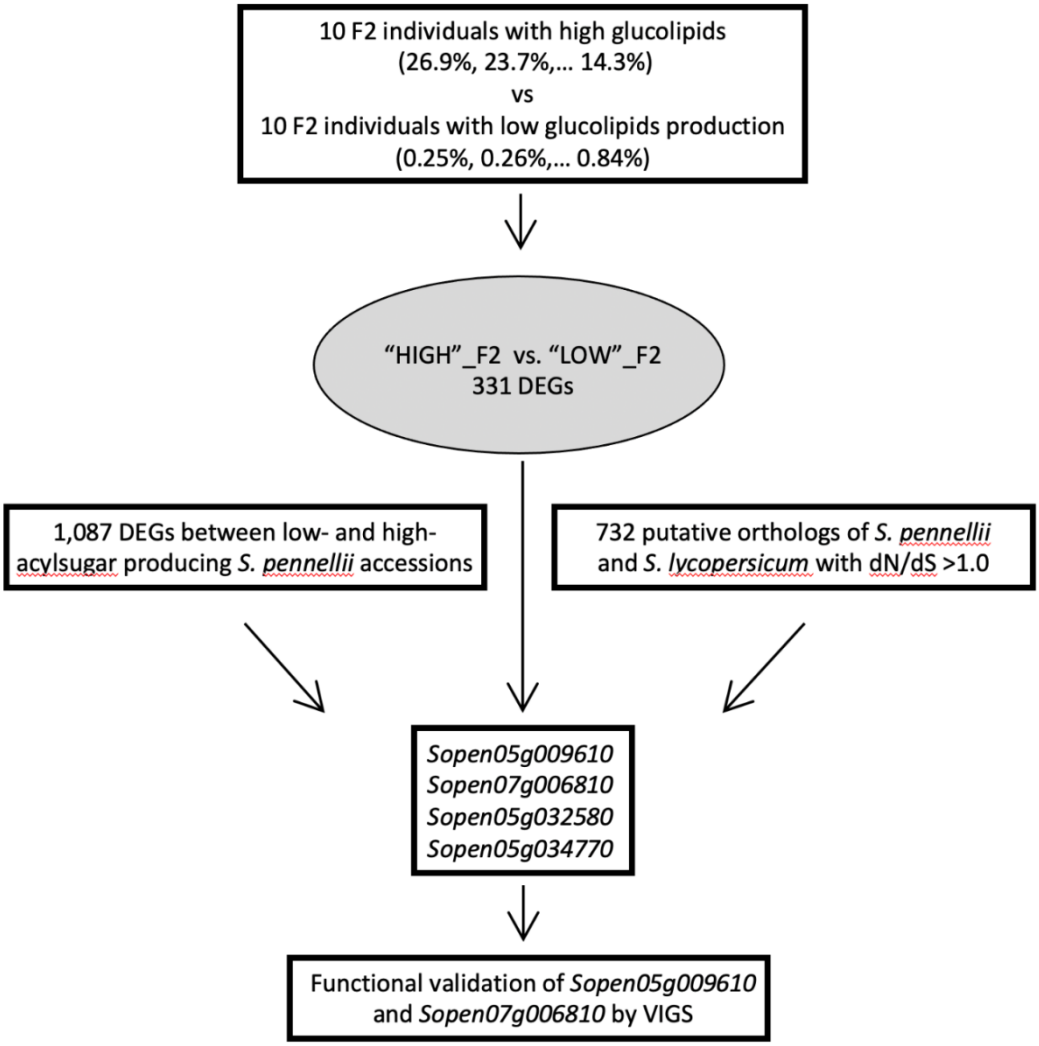

Figure 6. Workflow to identify candidate genes related to acylsugar metabolism. Comparative transcriptomics identified 331 DEGs from 10 high-acylsugar-producing F2 individuals with acylsugar amount between 26.90%-14.30% as "HIGH"-F2 group and top 10 low-acylsugar-producing F2 individuals with acylsugar amount between 0.25%-0.84% as "LOW"-F2 group. A genome-scale dN/dS ratio estimation was performed on *S. pennellii* and *S. lycopersicum* putative orthologs to identify 732 genes with dN/dS >1.0 to be under positive selection. We previously reported 1,087 DEG set between high-acylsugar–producing *S. pennellii* accessions and low-acylsugar–producing S. *pennellii* accessions (Mandal et al., 2020). In total, *Sopen05g009610*, *Sopen07g006810*, *Sopen05g032580* and *Sopen05g034770* were common in both DEG sets and were under positive selection. VIGS was applied to validate in vivo function of *Sopen05g009610* and *Sopen07g006810*.

Branched and straight fatty acyl chains are first synthesized in acylsugar biosynthetic network in *S. pennellii*, then acylated on the pyranose R2, R3, and R4 positions of sucrose (Leong et al., 2019; Fan et al., 2019). P-type acylsucroses were then cleaved to produce acylglucoses (Leong et al., 2019). We found genes involved in the synthesis of branched chain amino acids (BCAA), the precursors to branched chain fatty acids (BCFA), are upregulated in "HIGH"-F2 group comparing to "LOW"-F2 group (Figure 7, Table 3). Fatty acids synthesis genes, such as *Sopen12g004240* (3-ketoacyl-ACP synthase IV), *Sopen05g009610* (3-ketoacyl-ACP reductase I), *Sopen12g029240* (Enoyl-ACP reductase), *Sopen10g005030* (3-ketoacyl-CoA synthase I), were also upregulated in "HIGH"-F2 group (Figure 7, Table 3). Previously reported acyl-CoA synthetase (ACS) genes, which convert free medium chain fatty acids produced by de novo fatty acid biosynthesis to medium chain acyl-CoAs, as well as acylsucrose acyltransferases (ASATs) which utilize sucrose and acyl-CoA substrates to produce acylsucroses were found upregulated in "HIGH"-F2 group (Schilmiller et al., 2012; Schilmiller et al., 2015; Fan et al., 2015; Fan et al. 2020; Fan et al., 2017; Figure 7, Table 3). Three ATP-binding cassette (ABC) transporter family genes (*Sopen12g034820*, *Sopen04g005380*, and *Sopen03g001870*) were upregulated in "HIGH"-F2 group (Figure 7, Table 3). A special ribulose bisphosphate carboxylase (RUBISCO) small subunit, *Sopen07g006810*, required for CO2 fixation, was upregulated in "HIGH"-F2 group (Figure 7, Table 3), indicating specialized photosynthetic activities in glandular trichomes of *S. pennellii*. Transcriptomic comparisons between 10 high- and 10 low-glucolipid-producing F2 plants suggested the

expression levels of critical genes involved in fatty acids synthesis, acyl-activating process, central carbon metabolic and acylsugar export were positively correlated with acylsugar accumulation.

Expression profiles of two genes encoding UDP-Glc:FA GT and SCPL GAT, respectively (Ghangas and Steffens, 1993; Kuai et al., 1997; Li et al., 1999; Li and Steffens, 2000) did not positively correlate with acylsugar levels (Figure 7, Table 3).

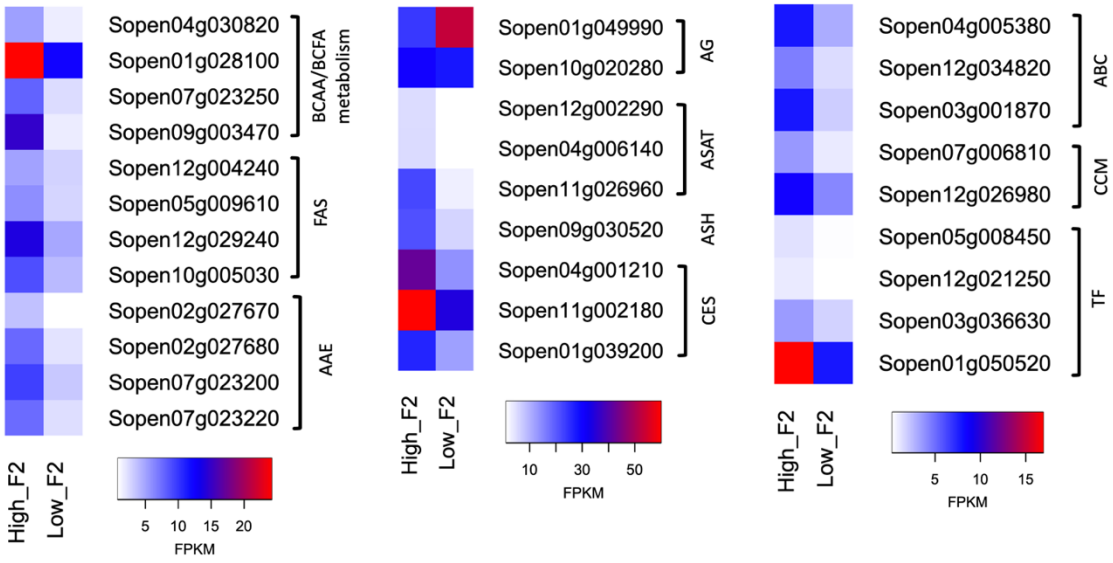

Figure 7. Expression levels of genes related or putatively related to acylsugar metabolism in the F2 population by heatmaps.
Genes are identified with annotated gene ID (Sopen IDs). AAE = acyl-activating enzyme; ABC = ABC transporters; AG = acylglucose previous model genes; ASAT = acylsucrose acyltransferase; ASH = acylsugar hydrolase; CES = Carboxylesterase; BCAA = branched-chain amino acid; BCFA = branched-chain fatty acids; CCM = central carbon metabolism; FAS = fatty acid synthase complex components; TF = Transcription factor; FPKM = fragments per kilobase of transcript per million mapped reads.

Table 3. Differentially expressed genes involved in acylsugar production identified from F2 population. 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 F2.

Gene ID	"HIGH"- vs. "LOW"-F2		Annotation
	Log_2FC	FDR	
Sopen09g003470	3.10	2.09E-02	Threonine dehydratase
Sopen04g030820	1.66	4.13E-02	Branched-chain aminotransferase-2
Sopen10g005030	1.15	5.97E-05	3-ketoacyl-CoA synthase 1
Sopen05g009610	1.15	1.20E-03	Beta-ketoacyl-ACP reductase
Sopen02g027670	2.14	6.74E-03	Acyl-activating enzyme 1
Sopen02g027680	1.88	1.88E-05	Acyl-activating enzyme 1
Sopen07g023200	1.51	1.40E-02	Acyl-activating enzyme 1
Sopen07g023220	1.69	2.46E-02	Acyl-activating enzyme 1
Sopen01g049990	-1.15	3.64E-02	UDP-glucose:fatty acid glucosyltransferase
Sopen10g020280	0.13	5.96E-01	Serine carboxypeptidase-like glucose acyltransferase
Sopen12g002290	2.21	2.58E-04	Acylsugar acyltransferase 1
Sopen04g006140	2.12	1.66E-02	Acylsugar acyltransferase 2
Sopen11g026960	2.87	4.86E-04	Acylsugar acyltransferase 3
Sopen09g030520	1.81	9.44E-07	Acylsugar acylhydrolase 3
Sopen01g039200	1.10	8.90E-04	Carboxylesterase
Sopen04g005380	1.38	3.94E-03	ABC transporter G family
Sopen12g034820	1.67	1.02E-02	Pleiotropic drug resistance protein 1-like (ABC-G family)
Sopen03g001870	2.01	6.01E-05	ABC transporter B family
Sopen07g006810	1.85	8.95E-03	RUBISCO small subunit
Sopen12g026980	1.01	2.46E-03	Glutathione S-transferase

Gene ID	"HIGH"- vs. "LOW"-F2		Annotation
	Log_2FC	FDR	
Sopen05g008450	1.73	7.81E-02	AP2 domain
Sopen12g021250	1.66	1.59E-02	AP2 domain
Sopen01g050520	1.12	3.57E-02	Myb domain

Genome-scale dN/dS ratio estimation of *S. pennellii* and *S. lycopersicum* putative orthologs

To further identify candidate genes, we performed a genome-scale dN/dS ratio (Yang and Bielawski, 2000) estimation of *S. pennellii* and *S. lycopersicum* putative orthologs to identify rapidly evolving genes. Putative *S. lycopersicum* orthologs of *S. pennellii* genes were identified by a reciprocal BLAST between all annotated *S. pennellii* and *S. lycopersicum* genes (Supplemental Data 5). A set of 19,984 reciprocal best Blast hit (RBH) orthologs pairs with E < 1e-10 was selected to estimate synonymous and nonsynonymous substitution rates. The yn00 maximum likelihood method from the PAML package (Yang et al., 1997) was used to calculate dN/dS ratios between putative *S. pennellii* and *S. lycopersicum* orthologs pairs (Supplemental Dataset 6: Sheet 1). This analysis yielded a genome-scale mean dN/dS ratio of 0.3273 (Figure 8). A total of 732 genes with dN/dS >1.0 were identified to be under positive selection (Supplemental Dataset 6: Sheet 2). In addition, the dN/dS ratios of 339 putative orthologs pairs were uncalculated because they all had dS=0. These genes were ignored in this study. Compared to *S. lycopersicum*, *S. pennellii* produces large amounts of acylsugar which may provide an advantage in the arid environment. Thus, we hypothesized that some

genes potentially associated with high abundance acylsugar production in *S. pennellii* were under positive selection and had dN/dS >1.0.

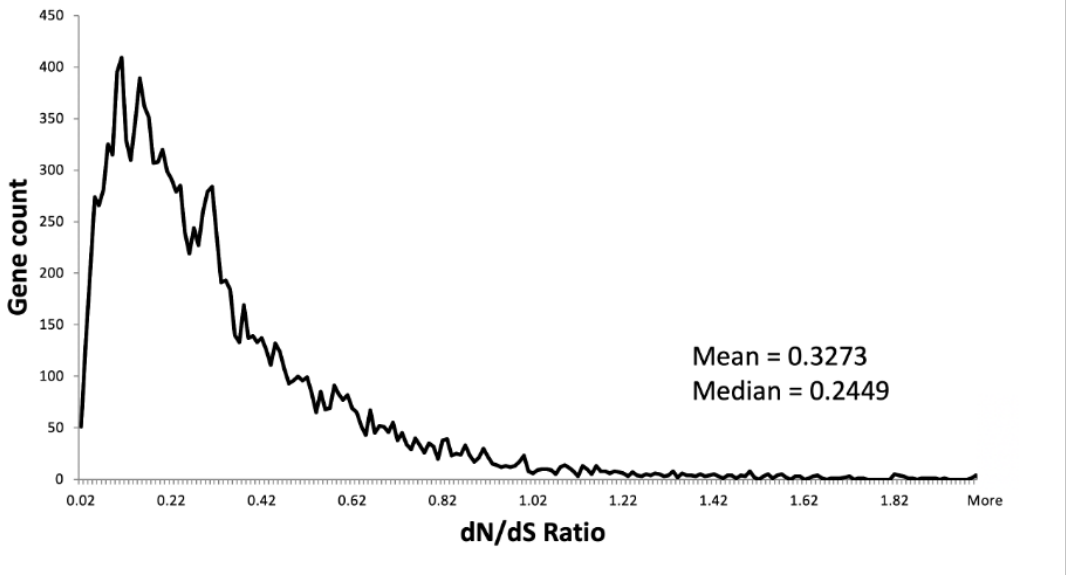

Figure 8. Histogram of dN/dS ratios of putative orthologous pairs from *S. pennellii* and *S. lycopersicum* calculated using the yn00 method in PAML program.

We previously reported 1,087 differentially expressed genes (DEGs) between high- and low-acylsugar–producing *S. pennellii* accessions (Mandal et al., 2020). To Identify genes that were common in both DEG sets and were under positive selection, we created a Venn diagram to calculate the intersections of three gene lists: DEGs in "HIGH"- and "LOW"- F2, DEGs in high- and low-acylsugar–producing *S. pennellii* accessions and genes with dN/dS>1 between *S. pennellii* and *S. lycopersicum* putative orthologs. *Sopen05g009610*, *Sopen07g006810*, *Sopen05g032580* and *Sopen05g034770* occurred in all three sets (Figure 9, Table 4).

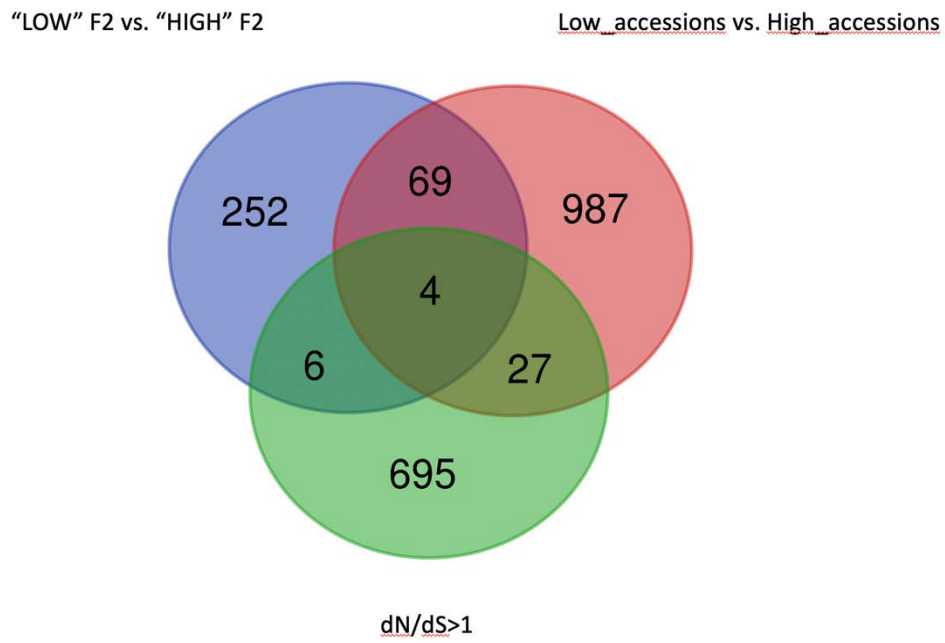

Figure 9. Venn diagram showing the intersections of three gene lists: DEGs in "HIGH"- vs "LOW"- F2, DEGs in high- and low-acylsugar–producing *S. pennellii* accessions and genes with dN/dS>1.0 between *S. pennellii* and *S. lycopersicum* putative orthologs. Numbers represent gene numbers in the corresponding intersection.

Sopen05g009610, annotated as a 3-ketoacyl-ACP reductase I, which is involved in fatty acid biosynthesis, had 2.2-fold higher expression in the "HIGH"-F2 group (FDR = 0.001), 10-fold higher expression in the high-acylsugar–producing *S. pennellii* accessions group (FDR = 3.80e-11), and a dN/dS = 2.80 (Table 4).

Sopen07g006810, annotated as a RUBISCO small subunit, had 3.6-fold higher expression in the "HIGH"-F2 group (FDR = 0.009), 32-fold higher expression in the high-acylsugar–producing *S. pennellii* accessions group (FDR = 6.70e-12), and a dN/dS = 2.91 (Table 4).

Table 4. Four genes common in both DEG sets and were under positive selection. dN/dS rank represents the corresponding rank of selected gene in 732 putative orthologs with dN/dS >1.0. Log_2FC indicates log_2 (fold-change). Positive and negative log_2FC values indicate higher and lower expression levels, respectively, in high-acylsugar–producing accessions or "HIGH"-F2 group. Full list is given in Supplemental Dataset 8.

Gene ID	dN/dS rank	dN/dS	high/low Log_2FC	HIGH/LOW-F2 Log_2FC	Annotation
Sopen07g006810	29	2.91	5.01	1.85	RuBisCO small subunit
Sopen05g009610	35	2.8	3.32	1.15	3-ketoacyl-ACP reductase
Sopen05g034770	199	1.59	-3.48	-1.89	uncharacterized protein LOC107020674
Sopen05g032580	227	1.52	2.97	2.1	Induced stolen tip protein TUB8-like

Sopen05g032580, annotated as a TUB8-like induced stolen tip protein, had 4.3-fold higher expression in the "HIGH"-F2 group (FDR = 0.013), 7.8-fold higher expression in the high-acylsugar–producing *S. pennellii* accessions group (FDR = 6.60e-05), and a dN/dS = 1.52 (Table 4).

Sopen05g034770 had 3.7-fold higher expression in the "LOW"-F2 group (FDR = 0.017), 11.2-fold higher expression in the low-acylsugar–producing *S. pennellii* accessions group (FDR = 2.30e-06), and dN/dS = 1.59 (Table 4). However, this gene was not annotated from published references. The sequences of *Sopen05g034770* were examined and no conserved domains have been identified from NCBI Conserved Domain Search (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi), and no signal peptide sequence could be determined by SignalP-5.0

(http://www.cbs.dtu.dk/services/SignalP/) for this predicted gene. The expression of this gene is inversely correlated with acylsugar production, and it was not analyzed further.

In vivo functional validation of *Sopen05g009610* and *Sopen07g006810*

Most genes reported to be involved in tomato acylsugar metabolism are preferentially expressed in trichomes relative to the underlying tissue (Schilmiller et al., 2012; Fan et al., 2015; Fan et al., 2020, Mandal et al., 2020). Published trichome RNA sequencing data indicated that expression of *Sopen05g009610, Sopen07g006810* and *Sopen05g032580* were trichome enriched in *S. pennellii*, while *Sopen05g034770* was not (Fan et al., 2020). We used quantitative reverse transcription PCR (qRT-PCR) to confirm the trichome enriched expression pattern of *Sopen05g009610* and *Sopen07g006810* (Figure 4). *Sopen05g009610* and *Sopen07g006810* have 110-and 200-fold, respectively, enrichment of transcript in *S. pennellii* LA0716 stems trichomes than underlying shaved stem tissue ($P < 0.001$, Welch two-sample t test, Figure 4). The qRT-PCR results confirmed that the *Sopen05g009610* and *Sopen07g006810* are highly expressed in *S. pennellii* LA0716 trichomes.

To pursue the hypothesis that *Sopen05g009610* (*SpKAR*) and *Sopen07g006810* (*SpRbcS*) are involved in acylsugar biosynthesis, virus-induced gene silencing (VIGS) was prepared to test *in vivo* roles for *Sopen05g009610* and *Sopen07g006810* in acylsugar production. These two genes were targeted in *S. pennellii* LA0716 for VIGS using tobacco rattle virus (TRV)-based silencing vectors (Dong et al., 2007), and the carotenoid biosynthetic enzyme, *S. pennellii* ortholog of phytoene desaturase (PDS), was

used as positive control (Figure 10). Extracts of trichome and leaf surface metabolites from these VIGS lines were analyzed by liquid chromatography–mass spectrometry (LC-MS). Total acylsugar levels decreased by 23% in *Sopen07g006810* VIGS group, compared to empty vector control group (Dunnett's test, $P < 0.05$, Figure 11). In contrast, we did not observe statistically significant change in total acylsugar level upon suppression of *Sopen05g009610* (Figure 11).

Figure 10. VIGS of a phytoene desaturase gene (*Sopen03g041530*), which was used as a visual positive control to evaluate the efficiency of VIGS and determine the harvest time.

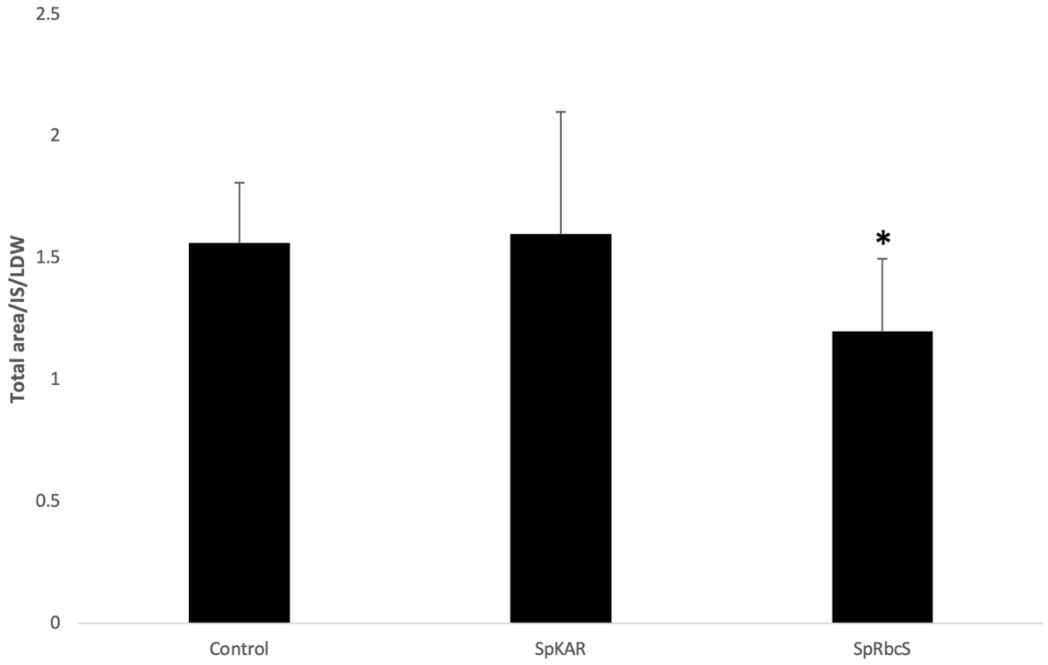

Figure 11. Acylsugar quantification by LC-MS. Extracts of trichome and leaf surface metabolites from control (empty TRV vector) and VIGS lines were analyzed by liquid chromatography–mass spectrometry (LC-MS). Peak areas were normalized by internal standard and leaf dry weight. Total acylsugar levels decreased by 23% in *Sopen07g006810*-VIGS group, compared to empty vector control. No statistically significant change of acylsugar level was observed upon suppression of *Sopen05g009610*. Welch's t test was used to compare the acylsugar abundance of *Sopen07g006810*/*Sopen05g009610*-VIGS and empty vector plants. Asterisk denote $P < 0.05$. Error bars indicate standard error (n = 10 individual plants).

Despite the lack of influence on total acylsugars, silencing the *Sopen05g009610* transcript led to 30% reduction in acylsugar medium-chain fatty acids production in C10 acyl chains amount (Dunnett's test, $P < 0.05$; Figure 12), which is consistent with its predicted role in medium-chain fatty acids biosynthesis. Suppression of *Sopen07g006810* transcript also resulted in 20% lower level of acylsugar C5 chain amount (Dunnett's test, $P < 0.05$; Figure 12). These numbers are likely to be

underrepresented since VIGS in *S. pennellii* is incomplete and inconsistent. Additionally, since no significant morphological differences were observed between silenced plants and control plants (Figure 13), better area-selective metabolite profiling was not possible. Nevertheless, transcript enrichment in trichomes of the two candidate genes, together with VIGS assay validation, indicate their impact on acylsugar biosynthesis.

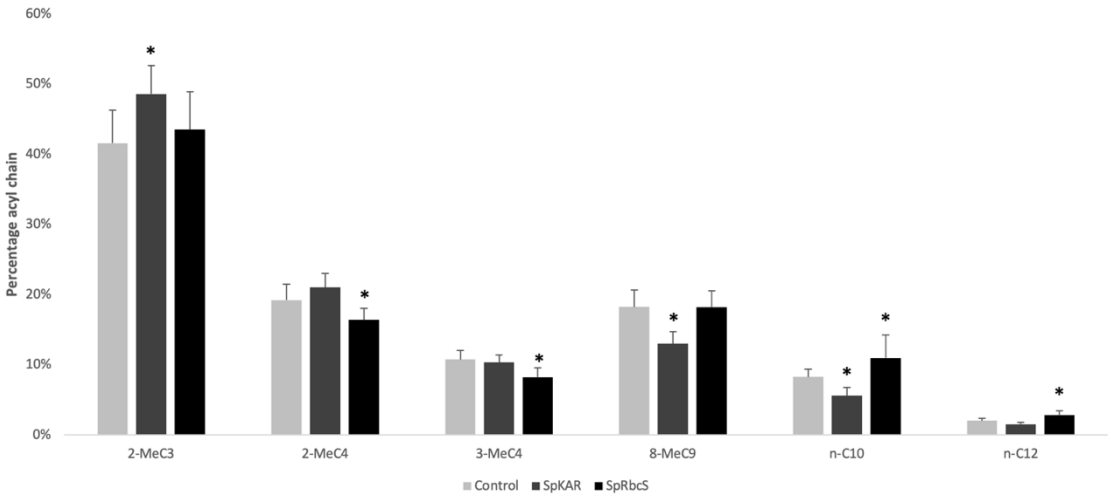

Figure 12. Acyl chain composition analysis by GC-MS. Acyl chain composition of acylsugars extracted from leaves of control plants (empty TRV vector) and VIGS targeted plants of *Sopen05g009610* and *Sopen07g006810*, respectively. Silencing of *Sopen05g009610* transcript resulted in 30% reduction in acylsugar SCFAs production in C10 acyl chains amount. Suppression of *Sopen07g006810* transcript resulted in 20% lower level of acylsugar C5 chain amount. Relative abundances of ethyl esters were determined by calculating the corresponding ethyl ester peak area over the total acyl chain peak area from GC-MS results. Dunnett's test was used to calculate statistical significance between *Sopen07g006810*/*Sopen05g009610*-VIGS and empty vector plants. Asterisk denote $P < 0.05$. Error bars indicate standard error (n = 10 individual plants).

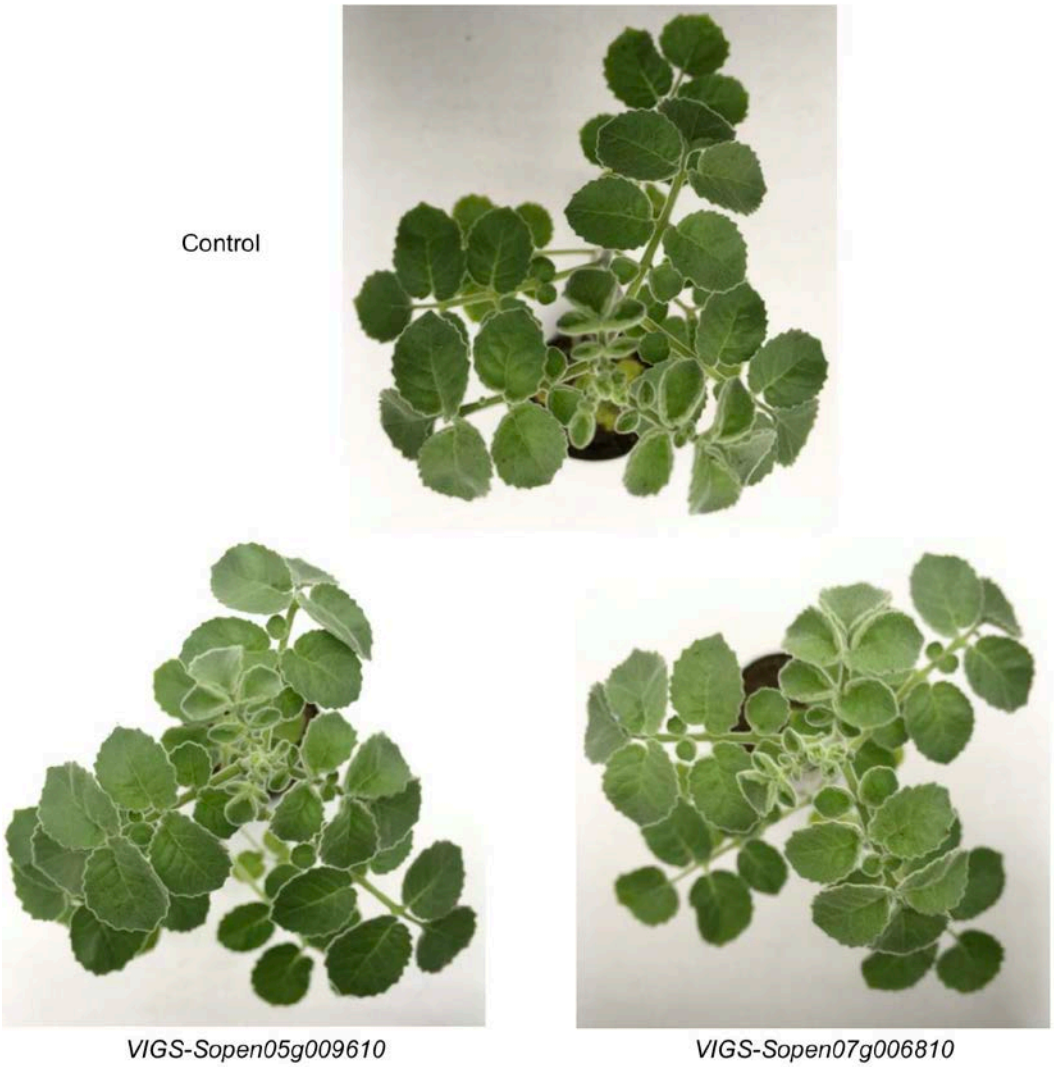

Figure 13. Morphology of *S. pennellii* plants belonging to the control group and VIGS groups. No significant differences were observed.

Phylogenetic analyses of 3-ketoacyl-ACP reductase and RUBISCO small subunit sequences

We performed phylogenetic analysis to reconstruct the evolutionary history of 3-ketoacyl-ACP reductase I and RUBISCO small subunit protein sequences among different solanaceous species (Figure 14, Figure 15). Phylogenetic positions of *Sopen05g009610* and *Sopen07g006810* indicated distinct secondary metabolism clades reside outside of the primary metabolism clades. This observation suggested duplication events leading to different clades of the phylogenetic tree related to evolution of acylsugar biosynthesis activities.

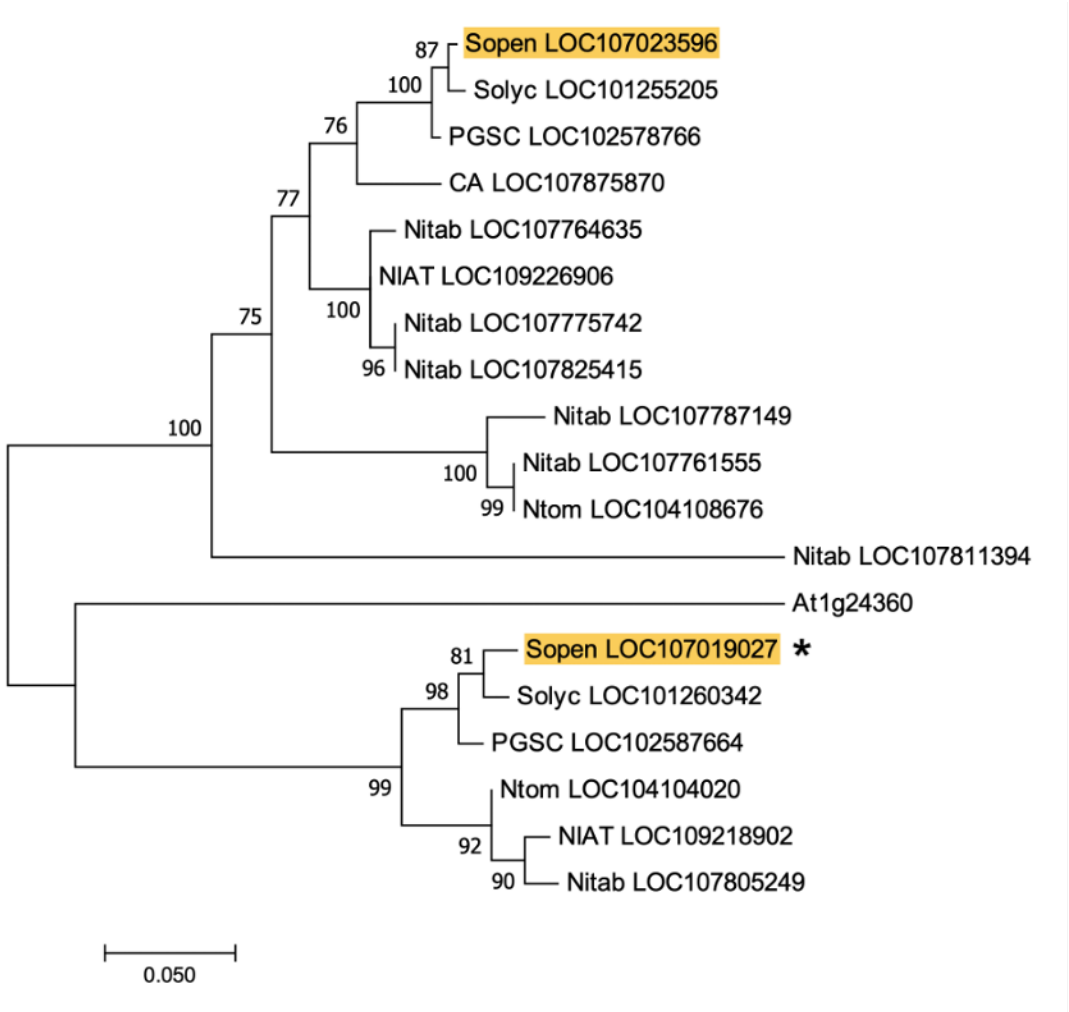

Figure 14. Phylogenetic analysis of SpKAR sequences.
Phylogenetic constructions of SpKAR sequences from different solanaceous species and *Arabidopsis thaliana*. Gene IDs are NCBI sequence IDs. Phylogenetic constructions were carried out using MEGA X (Kumar et al., 2018). Amino acid sequences were aligned with MUSCLE (Edgar, 2004) under default parameters. A maximum likelihood method based on Jones-Taylor-Thornton (JTT)+G model (Jones et al., 1992) was used to generate the phylogenetic trees, with 1000 bootstrap replicates. Phylogenetic trees were drawn to scale, with branch lengths computed using the Poisson correction method (Zuckerkandl and Pauling, 1965) in the number of amino acid substitutions per site. SpKAR sequences from *Solanum pennellii* are indicated with yellow highlights. *Sopen05g009610* is indicated by an asterisk.
Sopen= *Solanum pennellii*, Solyc= *Solanum lycopersicum*, PGSC= Potato (*Solanum tuberosum*), CA=*Capsicum annuum*, Nitab= *Nicotiana tabacum*, NIAT= *Nicotiana attenuata*, Ntom= *Nicotiana tomentosiformis*, At= *Arabidopsis thaliana*

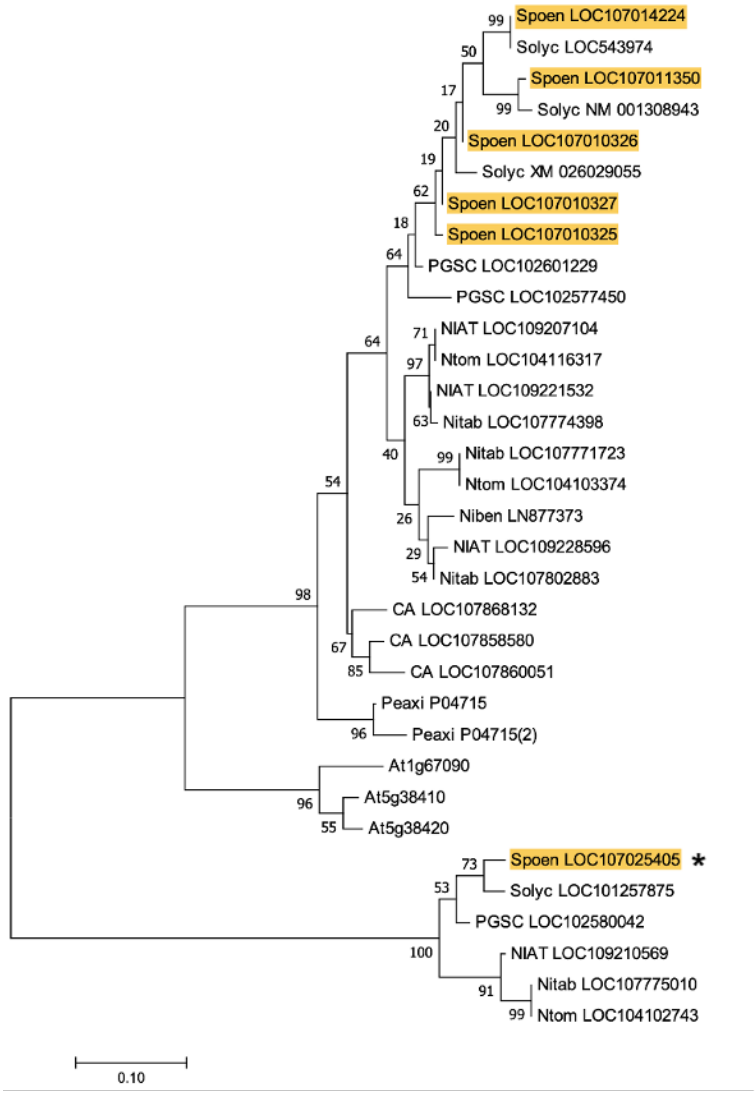

Figure 15. Phylogenetic analysis of SpRbcS sequences.
Phylogenetic constructions of SpRbcS sequences from different solanaceous species and *Arabidopsis thaliana*. Gene IDs are NCBI sequence IDs. Phylogenetic constructions were carried out using MEGA X (Kumar et al., 2018). Amino acid sequences were aligned with MUSCLE (Edgar, 2004) under default parameters. A maximum likelihood method based on Jones-Taylor-Thornton (JTT)+G model (Jones et al., 1992) was used to generate the phylogenetic trees, with 1000 bootstrap replicates.
SpRbcS sequences from *Solanum pennellii* are indicated with yellow highlights. *Sopen07g006810* is indicated by an asterisk.
Sopen= *Solanum pennellii*, Solyc= *Solanum lycopersicum*, PGSC= Potato (*Solanum tuberosum*), CA=*Capsicum annuum*, Nitab= *Nicotiana tabacum*, NIAT= *Nicotiana attenuata*, Peaxi= *Petunia x hybrida*, Ntom= *Nicotiana tomentosiformis*, At= *Arabidopsis thaliana*

CHAPTER III

DISCUSSION

Acylsugars are a group of specialized metabolites with high viscosity that have been reported in six plant families (Wu et al., 2013; Liu et al., 2019). Due to the insecticidal and anti-pathogen properties of these compounds, increasing acylsugars production has been a long-time goal in tomato breeding strategies (Bonierbale et al., 1994; Lawson et al., 1997). In *S. pennellii*, a wild relative of *S. lycopersicum*, acylsugars constitute a particularly larger proportion of leaf biomass than cultivated tomato (Fobes et al., 1985). While many steps of acylsugars synthesis pathways have been identified (reviewed by Fan et al., 2019), the mechanism prompting acylsugar accumulation to such a high level in *S. pennellii* remains unknown. In this study, we used comparative genomic approaches in two independent experiments: a differential gene expression analysis between high- and low-acylsugar-producing *S. pennellii* accessions group, and another differential gene expression analysis between 10 high- and 10 low-acylsugar-producing F2 plants of *S. pennellii* LA 0716 and *S. lycopersicum* cv. VF36 cross, to further investigate the biosynthesis of trichome acylsugar in *S. pennellii*. We confirmed two candidate genes, *Sopen05g009610* and *Sopen07g006810*, through VIGS, to be involved in medium-chain fatty acids synthesis pathway and carbon fixation associated with acylsugar production, respectively. These results provide a strong support for the involvement of novel candidate genes we have identified in acylsugar biosynthesis, and

validate our approach of combining differential gene expression and evolutionary analysis.

Acylsugar metabolic genes

Our initial hypothesis for this project was that expression of genes previously reported to be involved in acylsugar metabolism and expression of unknown genes required for acylsugar production would correlate with acylsugar levels in different accessions of *S. pennellii*. Our approach was validated by the identification of several DEGs with known functions in acylsugar biosynthesis, such as BCAA metabolic genes and acylsucrose biosynthetic genes (ASATs). We also treated *S. pennellii* leaves with imazapyr to inhibit BCAA biosynthesis and then analyzed the expression profiles before and after the treatment to refine the list of acylsugar candidate genes. Our results showed that genes related to SCFA metabolism and other candidate genes of acylsugar metabolism responded to imazapyr treatment in a dose-dependent manner. Repression of acylsugar metabolic genes after imazapyr treatment in this study is distinct from the induction of general stress response genes that can result from amino acid starvation (Zhao et al., 1998).

We found that genes involved in both biosynthesis of BCAAs and BCFAs were upregulated in high-acylsugar-producing accessions (Table 2), which was consistent with previous feeding studies that showed acylsugar BCFAs are derived from BCAAs (Walters and Steffens, 1990). We also identified putative SCFAs biosynthetic genes and acyl-activating enzymes (AAE1) that may provide SCFA-CoA molecules for ASATs in

acylsucrose biosynthesis based on not only their strong positive correlation with acylsugar accumulation (Figure 3), but also significant downregulation of these candidate genes in response to imazapyr treatment (Supplemental Dataset 2: Sheet 2). *Sopen11g007710*, which is predicted to encode an Mpv17/PMP22 family peroxisomal membrane protein, is presumably involved in controlling permeability of peroxisomal membrane (Brosius et al., 2002), and this gene showed strong response to imazapyr, further supporting a critical role of the peroxisome in acylsugar metabolism. It remains unknown how acylsugars are secreted out of trichome cells.

Here, we identified three putative ABC transporter genes as candidates for acylsugar transport based on their expression profiles in the comparative genomic analysis as well as their response to imazapyr treatment (Figure 3, Supplemental Dataset 2: Sheet 2). We also identified DEGs involved in central carbon metabolism, such as genes encoding RUBISCO small subunit and NADP-malic enzyme. These enzymes may provide carbon in *S. pennellii* trichomes to support acylsugar biosynthesis. Upregulation of genes encoding starch synthase and a monosaccharide transporter in high-acylsugar-producing accessions (Table 2) is consistent with the proposal that glandular trichomes import most of their fixed carbon from underlying leaf tissues to support acylsugar biosynthesis, despite being able to perform some photosynthetic activities (Balcke et al., 2017). Furthermore, significant downregulation of these DEGs in response to imazapyr (all but the starch synthase gene, Supplemental Dataset 2: Sheet 2) further supports their involvement in acylsugar biosynthesis. I also performed qRT-PCR to investigate

expression levels of selected candidate genes in isolated stem trichomes versus shaved stems, and found that they are preferentially expressed in trichomes (Figure 4).

After obtaining the above acylsugar related candidate biosynthetic and regulatory gene pool which includes 1,087 DEGs through comparative transcriptomics of different acylsugar producing accessions of *S. pennellii*, we also observed a phenotypic segregation of total acylsugar amounts produced by the F2 progeny of *S. pennellii* LA 0716 and *S. lycopersicum* cv. VF36 cross. Thus, it was possible to further refine the list of candidate genes by conducting transcriptomic comparation between "HIGH"- and "LOW"-F2 accessions. We identified 331 DEGs across 10 high- and 10 low-acylsugar-producing F2 individuals. These DEGs included key genes involved in acylsugar biosynthesis such as BCAA/BCFA/SCFA metabolic genes, ASATs and ACSs (Figure 7, Table 3). This number is less than the 1,087 DEGs data set identified in a previous study (Mandal et al., 2020), which can be explained from two angles. First, due to the genetic background of F2 population, the number of selected F2 individuals and the phenotypic differences within each group, there was more genetic variation within the "HIGH"- or "LOW"- F2 group than within high- or low- acylsugar producing *S. pennellii* accessions group. Second, it is necessary for the "HIGH"- F2 individuals to acquire all the core genes controlling high acylsugar producing phenotype from *S. pennellii* parent to accumulate high level of acylsugars, whereas "LOW"- F2 individuals might each have different alleles for any one or more of these core genes with the others remaining active and highly expressed. Meanwhile, one large group of genes present in the previous study (Mandal et al., 2020), but not found in the present F2 analysis are defense-related genes.

Genome-scale dN/dS calculation

To further identify candidate genes associated with acylsugar production, I performed a genome-scale dN/dS ratio estimation of *S. pennellii* and *S. lycopersicum* putative orthologs. Because high acylsugar levels provide *S. pennellii* advantages toward adaptation to arid environment and resistance to insect herbivores (Fobes et al., 1985; Puterka et al., 2003; Alba et al., 2009; Leckie et al., 2012), I hypothesized that genes related to the high acylsugar phenotype might be under positive selection, and these genes may have dN/dS ratios greater than 1. Although there are several possible explanations for high dN/dS ratio other than positive selection, such as intrinsic nonlinearity of mutations accumulation, relaxation of negative selection and genetic drift (Lahti et al. 2009; Montoya-Burgos 2011; Strasburg et al. 2011; Mugal et al. 2014; Hawkins et al., 2017), dN/dS ratio remains a good exploratory strategy for aiding in the identification of candidate genes under positive selection. I identified 732 genes with dN/dS >1.0 in this genome-wide calculation (Supplemental Dataset 6: Sheet 2). Four genes were in the intersection of both DEG sets and dN/dS >1 set, and I selected *Sopen05g009610* and *Sopen07g006810* for further *in vivo* validation. Expression of one of these four genes, *Sopen05g034770* (annotated as encoding a hypothetical protein), was inversely correlated to acylsugar accumulation. Its low expression level in LA0716 (FPKM = 0.90), made it a poor candidate for VIGS analysis so I did not study it further.

SpKAR-T Function

In addition to its key role in long chain fatty acid biosynthesis in plants, the FAS complex system also appears to be employed for medium-chain fatty acid synthesis, which contributes to acylsugar accumulation in tomato (van der Hoeven and Steffens, 2000; Kroumova and Wagner, 2003). Several genes with trichome-enriched expression, such as those encoding ASATs that add fatty acid chains to sugar backbone to yield acylsugar (Schilmiller et al., 2012; Schilmiller et al., 2015; Fan et al., 2016; Fan et al., 2017; Fan et al., 2019). However, few genes associated with FAS involved in BCFAs and SCFAs elongation have been identified, and the medium-acylsugar chain synthesis pathway has not been characterized in *S. pennellii*. We previously identified two trichome-enriched KAS enzymes of the FAS complex responsible for medium-chain fatty acid synthesis in *S. pennellii* (Mandal et al., 2020). In this study I identified *Sopen05g009610*, a trichome preferentially expressed FAS component 3-ketoacyl-ACP reductase (SpKAR-T), that affects medium-chain fatty acids synthesis pathway using genomic tools and functional validation. I noticed this gene since it was upregulated in "HIGH"- F2 group as well as in high- acylsugar–producing *S. pennellii* accessions and has an extreme high dN/dS ratio of 2.80. The role of SpKAR-T was tested further by VIGS in *S. pennellii*. Although suppression of *Sopen05g009610* did not have a significant effect on total acylsugar levels, its down-regulation did cause a reduction in 8-MeC9 and n-C10 output by 30% (Figure 12). This knockdown phenotype suggests that SpKAR-T contributes to medium-BCFA and SCFA production in *S. pennellii* trichomes. Other than the decrease of medium-chain fatty acid accumulation, we also observed an

increase in short-chain fatty acid production in acylsugars, suggesting that a supply shortage of FAS-mediated medium-chain fatty acid could result in a compensatory increase of short acyl chains on acylsugars. These data, together with the previously reported KAS enzymes (Mandal et al., 2020), suggest involvement of plastid FAS complex in medium-chain fatty acids biosynthesis in *S. pennellii*, which may be helpful for manipulating acylsugar fatty acids chain length in engineering tomato breeding lines with improved insect resistance.

SpRbcS-T Function

In plants, photosynthesis predominantly takes place in mesophyll cells of leaves. However, photosynthetic activities are observed in tomato and tobacco trichomes secretory cells as well (Kandra and Wagner, 1988; Pike and Howells, 2002; Balcke et al., 2017), though the levels of photosynthetic activities in trichomes were significantly lower than in the rest of the leaf (Laterre et al., 2017; Balcke et al., 2017). Glandular trichomes synthesize and secrete high volume of secondary metabolites, such as acylsugars in *S. pennellii*, which require large amounts of photosynthate imported from other tissues due to insufficient photosynthesis in trichomes (Balcke et al., 2017). Photosynthetic activities in trichomes were previously considered to be similar to primary photosynthesis in other leaf cells, and little information has been reported concerning the biochemical and physiological roles of photosynthesis and RUBISCO in trichomes. Recently, a tobacco trichome RUBISCO S-subunit (NtRbcS-T) was reported with lower affinity for CO_2, and a higher catalytic activity. These changes decrease

carboxylation efficiency and increase acidic pH-dependent activity relative to RUBISCO S-subunit isoforms expressed in mesophyll cells, conferring C4-like kinetic properties to trichome RUBISCO (Laterre et al., 2017; Martin-Avila et al., 2020).

In my transcriptome comparison and dN/dS analysis I found *Sopen07g006810* encoding a trichome-enriched RUBISCO small subunit (SpRbcS-T) in *S. pennellii* as one of the up-regulated genes in both DEG sets and in the dN/dS>1 set (Table 4). Total acylsugar levels were reduced by 23% in *S. pennellii* after VIGS suppression of SpRbcS-T (Figure 11). I also observed that there was a 20% drop in the level of short acylsugar chains upon the suppression of *Sopen07g006810* transcript (Figure 12).

Large amounts of CO_2 are generated within trichomes by very active specialized metabolic pathways, such as those producing pathogen-deterring diterpenes and acylsugars (Wagner et al., 2004; Laterre et al., 2017; Pottier et al., 2018). For instance, CO_2 is produced during the first step in the synthesis of BCAAs catalyzed by acetolactate synthase (ALS), and during conversion of 3-isopropylmalate into 4-methyl-2-oxopentanoate catalyzed by 3-isopropylmalate dehydrogenase (IPMD) (Figure 2A, Page 20). Accumulation of CO_2 might inhibit these reactions towards the branched-chain amino acids synthesis, resulting in a decrease of branched acyl chains level on acylsugars.

The presumed high cellular CO_2 concentration in trichomes is similar to C4 plants, and might result in an acidic pH environment, since the thick cell wall and cuticle of trichomes may lower gas exchange with the atmosphere (Kandra, L. and Wagner, J. 1988; Savchenko et al., 2000). In tobacco, NtRbcS-T with lower affinity for CO_2 and

preserved activity at more acidic pH might be involved in recycling metabolic CO_2 released during the specialized metabolism, thereby increasing the productivity and carbon efficiency of the trichome cells (Laterre et al., 2017; Martin-Avila et al., 2020).

Phylogenetic analysis of RUBISCO small subunit sequences revealed that SpRbcS-T sequence is clearly distinct from other RbcS isoforms encoded in the genome (Figure 15). This result is consistent with that previously published for NtRbcS-T, which belongs to a distinct phylogenetic cluster T (trichome) from the other RbcS sequences in another cluster M (mesophyll) (Laterre et al., 2017). Similarly, SpKAR-T segregates in a distinct lineage other than the KAR involved in primary metabolism in *S. pennellii* and other acylsugar-producing species (Figure 14). Our phylogenetic analysis of SpRbcS-T and SpKAR-T reflects another example of distinct specialized metabolism subclades different from the primary metabolism subclades, supporting the hypothesis that specialized metabolism requirements of trichome lead to the formation of neo-secondary metabolic enzymes through independent gene duplication and neo-functionalization of primary metabolic enzymes as evolutionary adaptation (Milo and Last, 2012; Moghe and Last, 2015; Ning et al., 2015; Laterre et al., 2017; Pottier et al., 2018; Maeda, 2019).

Specialized metabolism pathways and enzymes attract increasing attention in recent research, especially in studies regarding how plants evolved to mediate environmental interactions. In this work I used traditional genetic methods together with robust bioinformatics and functional genomics screening tools to extend our understanding of the genetic and biochemical mechanisms leading to acylsugar production in *S. pennellii*. Given the important roles of acylsugar in plant-defense

activities, full understanding of the acylsugar biosynthesis pathway will be essential for crop engineering to create pests resistant tomato breeding lines with improved acylsugar levels and compositions.

CHAPTER IV

MATERIALS AND METHODS

Plant growth conditions, generation of F2 cross progeny and acylsugar collection

Seeds of all wild tomato *Solanum pennellii* accessions and F1 generation (LA4135) of *Solanum pennellii* LA0716 *X Solanum lycopersicum* VF36 were obtained from the C.M. Rick Tomato Genetics Resource Center (University of California, Davis, http://tgrc.ucdavis.edu/). The LA4135 was self-pollinated for the construction of F2 progeny, resulting in 114 lines. Seeds were treated with 20%(v/v) bleach for 20 min and rinsed with deionized water three times before placing on moist filter paper in petri dishes. After germination, seedlings were transplanted to soil and grown in a growth chamber (24°C day/20°C night temperature, 150 µMol m^{-2} s^{-1} photosynthetically active radiation, 16-hour photoperiod, and 75% relative humidity). Secreted acylsugars were collected from three branches of 10-week-old F2 individuals as three replicates by dipping young leaves in ethanol for 3 seconds. Ethanol was completely removed by evaporation until dryness in fume hood. Dry weights of the extracted leaves were measured after drying in a 70°C oven for one week. Acylsugar abundance was calculated as a proportion of leaf dry weight. Raw data are available in Supplemental Dataset 3.

RNA extraction and sequencing

Three biological replicates were used for each accession in the "HIGH" vs. "LOW" comparison, and for the "MEDIUM" accession LA1302. Four biological replicates were used for each accession in the LA1920 vs. LA0716 comparison. Ten highest-acylsugar-producing F2 individuals (acylsugar amount measured as 26.86%, 23.67%, 20.82%, 18.60%, 17.41%, 17.20%, 17.19%, 16.22%, 14.93% and 14.35% of leaves dry weight, respectively) and 10 lowest-acylsugar-producing F2 individuals (acylsugar amount measured as 0.25%, 0.26%, 0.35%, 0.52%, 0.53%, 0.62%, 0.68%, 0.73%, 0.75% and 0.84% of leaves dry weight, respectively) were used in the "HIGH"- vs. "LOW"-F2 comparison. Before RNA extraction, young leaves from 10-week–old plants were dipped in ethanol for 3 seconds to remove the acylsugar from the leaf surface. Leaves were then immediately frozen with liquid nitrogen, and stored at -80°C until further use. Total RNA was isolated from leaves using the RNAqueous Total RNA Isolation Kit (Thermo Fisher Scientific), followed by the genomic DNA digestion using the TURBO DNA-Free Kit (Thermo Fisher Scientific). Quality of each RNA sample was assessed using the Agilent 2200 TapeStation software A01.04 (Agilent Technologies). In high- vs. low-acylsugar-producing accessions comparison experiments, RNA-Seq libraries were prepared from polyA$^+$-selected RNA samples using the TruSeq RNA Library Preparation kit v2 (Illumina), and then sequenced on the Illumina HiSeq 2500 v4 125 X 125-bp paired-end sequencing platform (High Output Mode) following manufacturer's specifications. Sequencing was performed at Texas A & M Genomics and Bioinformatics Service Center, College Station. In "HIGH- vs. "LOW"-acylsugar-

producing F2 comparison experiments, RNA sequencing libraries of 20 polyA^{+}-selected RNA samples were prepared using TruSeq Stranded mRNA Library Preparation Kit LT (Illumina). After quality control and quantitation, libraries were sequenced on the HiSeq 4000 (Illumina) 150-bp X 150-bp paired-end sequencing platform according to the manufacturer's specifications at the Texas A&M Genomics and Bioinformatics Service Center, College Station. Base calling was done by Illumina Real Time Analysis (RTA) v1.18.66.3 software in real time, and the resulting sequencer.bcl basecall files were demultiplexed and converted into FastQ files by Illumina Bcl2fastq v2.17.1.14 script conFigureBclToFastq.pl.

Differential expression analyses

Phred quality score distributions of sequencing reads were analyzed with the software FastQC v0.11.4 (Andrews, 2010). All RNA-Seq libraries had average Phred quality scores of more than 35.

In high- vs. low-acylsugar-producing accessions comparison experiments, approximately 30 million 125-bp paired-end reads were generated for each submitted RNA sample. RNA-seq data was processed using the pipeline from as previously described (Mandal et al., 2020). Lists of DEGs as determined by edgeR (Robinson et al., 2010) are given in Supplemental Dataset 1.

In "HIGH"- vs. "LOW"-acylsugar-producing F2 comparison experiments, approximately 31 to 48 million (36 million in average) 150-bp paired-end reads were generated for each submitted RNA sample. Paired-end reads were processed for quality

control using Trimmomatic v0.32 (Bolger et al., 2014b) with the following settings applied: ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10, LEADING = 20, TRAILING = 20, SLIDINGWINDOW = 4:20, MINLEN = 100. Approximately 68% of the reads in each data set passed the trimming filter. The trimmed reads were then mapped to the *S. pennellii* genome v2.0 (LA0716; Bolger et al., 2014a) using TopHat2 v2.1.0 (Kim et al., 2013) with the following parameters: -mate-inner-dist = 0, -mate-std-dev = 50, -read-realign-edit-dist = 1000, -read-edit-dist = 2, -library-type = fr-firststrand, -read-mismatches = 2, -min- anchor-len = 8, -splice-mismatches = 0, -min-intron-length = 50, -max- intron-length = 50,000, -max-insertion-length = 3, -max-deletion-length = 3, -max-multihits = 20, -min-segment-intron = 50, -max-segment-intron = 50,000, -segment-mismatches = 2, -segment-length = 25. Seventy percent to 85% of the trimmed reads were map to the *S. pennellii* genome successfully. Reads mapped to selected loci of interest were visualized with IGV (Thorvaldsdóttir et al., 2013). Aligned reads from TopHat2 were counted for each gene using HTseq package version 0.6.1 (Anders et al., 2015) with the following parameters: -f bam, -r name, -s reverse, -m union, -a 20. The count files were used to identify DEGs using edgeR package version 3.32.1 (Robinson et al., 2010). Fragments per kilobase per million mapped reads (FPKM) value for each gene in each sample was called with rpkm command in edgeR program. Genes with more than 1 count per million (CPM) in at least two samples were used for DEGs analysis. Library sizes were recalculated after the filtering process. DEGs were identified when P, corrected for multiple testing, was less than 0.05 (FDR < 0.05, Benjamini and

Hochberg, 1995), and fold change was greater than 2. Lists of DEGs as determined by edgeR (Robinson et al., 2010) are given in Supplemental Dataset 4.

Determination of putative orthologs and dN/dS estimation

Reciprocal BLAST was performed to determine putative orthology between *S. pennellii* genes and *S. lycopersicum* genes. We performed an all-versus-all BLAST between all annotated proteins of *S. pennellii* v2.0 (Bolger et al., 2014a) and *S. lycopersicum* ITAG3.0 (Tomato Genome Consortium, 2012; Fernandez-Pozo et al., 2015) with the following settings applied: minimum percentage identity = 70.0, minimum percentage query coverage = 50.0. Putative orthologs were identified from pairs of reciprocal best hits with the lowest e-value for each other. Genome-wide dN/dS were calculated between reciprocal best hits using yn00 maximum likelihood method in the PAML package (Yang 1997, Yang and Nielsen 2000).

Quantitative RT-PCR

Total RNA was harvested from stem trichomes and shaved stems of 10-week–old *S. pennellii* LA0716 plants as follows: stems were cut into small pieces and flash-frozen in liquid nitrogen, and trichomes were scraping off the frozen stems with a scalpel. For each sample, total RNA was isolated using the RNAqueous Total RNA Isolation kit (Thermo Fisher Scientific), followed by genomic DNA digestion with the TURBO DNA-free kit (Thermo Fisher Scientific). Reverse transcription was performed with 1 µg of total RNA to synthesize cDNA using SuperScript IV VILO Master Mix (Thermo

Fisher Scientific) according to the manufacturer's guidelines. Quantitative PCR (qPCR) was carried out with SYBR Green Master Mix (Bio-Rad) using QuantStudio 6 Flex Real-Time PCR System with the following temperature cycling conditions: 50°C for 2 min, 95°C for 10 min, followed by 40 cycles of 95°C for 15 s and 60°C for 1 min. SpASAT1_F/R, AP2_F/R, SpRbcS_F/R, SpKAR_F/R, KAS III_F/R, KAS IV/II-like_F/R, AAE_1_F/R, AAE_2_F/R, AAE_3_F/R, AAE_4_F/R, PMP22_F/R, ABC_1_F/R, ABC_2_F/R, ABC_3_F/R, and Ubiquitin_F/R primers were used to detect SpASAT1, AP2, SpRbcS, SpKAR, KAS III, KAS IV/II-like, AAE_1, AAE_2, AAE_3, AAE_4, PMP22, ABC_1, ABC_2, ABC_3, and Ubiquitin transcripts, respectively (Supplemental Dataset 7). The transcript levels for SpASAT1, AP2, SpRbcS, SpKAR, KAS III, KAS IV/II-like, AAE_1, AAE_2, AAE_3, AAE_4, PMP22, ABC_1, ABC_2 and ABC_3 were normalized to ubiquitin gene *Sopen02g027600* as the endogenous control. The comparative C_T ($\Delta\Delta C_T$) method was used to evaluate the quantitative variation of gene expression.

Virus-induced gene silencing

VIGS was performed using the tobacco rattle virus (TRV) vector system (Dong et al., 2007) and VIGS constructs were designed with the SGN online tool (http://vigs.solgenomics.net/). Sequences targeting fragments of SpKAR, SpRbcS, and the phytoene desaturase (PDS) gene were amplified with the primers listed in Supplemental Dataset 7, followed by cloning into pTRV2-LIC vector and transforming into *Agrobacterium tumefaciens* strain GV3101. *A. tumefaciens* GV3101 harboring pTRV1, pTRV2

constructs, or empty pTRV2 were grown overnight with 50 mg/mL kanamycin and 10 mg/mL gentamicin at 28°C. The cultures were centrifuged at 8,000g for 5 min at 4°C. The cells were washed and resuspended in infiltration buffer (10 mM MES pH 5.5, 10 mM $MgCl_2$, and 200 μM of acetosyringone). Cell suspensions were incubated at room temperature for 3 h, and then different pTRV2 constructs cultures were mixed with equal volume of pTRV1 cultures to reach the final OD600nm = 1 before infiltration. Individual *S. pennellii* LA0716 seedlings were inoculated at the first true leaf stage. Plants were incubated in the growth chamber with conditions mentioned before for approximately six weeks. The PDS gene *Sopen03g041530* was used as a visual positive control to evaluate the efficiency of VIGS and determine the harvest time (Figure 10). At least ten plants were analyzed for each construct.

Acylsugar quantification by LC-MS

For the analysis of acylsugar abundance, acylsugars on leaf surface were extracted from similar-sized young leaflets by submerging them in 10 ml of acetonitrile/iso-propanol/water (3:3:2, v/v/v) mixture extraction solvent containing 0.1% formic acid and 10 μM propyl 4-hydroxybenzoate as internal standard, followed by gentle mixing for 2 min. The extraction solvent was stored at −20°C for further use. Dry weights of the extracted leaflets were measured after one week drying in a 70°C oven. Peak areas were normalized by internal standard and leaf dry weight. Welch's t test was used to compare the acylsugar abundance of SpKAR/SpRbcS-VIGS and empty vector plants.

Acylsugar acyl chain composition analysis by GC-MS

For the analysis of acyl chains on acylsugars, Leaf surface acylsugars were extracted from similar-sized young leaflets by submerging them in ethanol for 3 seconds, followed by the evaporation of ethanol to complete dryness and re-dissolving in n-heptane. For the transesterification reaction of acylsugar extractions, 500 μL of 20%(v/v) sodium ethoxide in ethanol was added to 1 mL of heptane mixture and incubated for 10 min at room temperature, followed by gentle mixing. The heptane layer was removed and washed with 500 μL of saturated sodium chloride in water by mixing to facilitate a phase separation, repeated twice. The final top heptane layer containing the fatty acid ethyl esters was transferred to a 2 mL glass vial for gas chromatography-mass spectrometry (GC-MS) analysis using Trace GC Ultra with DSQII system (Thermo Fisher Scientific). A 30-m, 0.25-mm fused silica column with a 0.25-μm film thickness stationary phase (DB-5MS; Agilent) was used as the GC column. Injection of 1 μL of heptane extract was performed using split injection (1:50). The following gas chromatography program was performed: injector temperature, 225°C; starting column temperature, 30°C held for 2 min; ramped at 15°C/min until 300°C; and then held for 5 min. The helium carrier gas flow was 1.5 mL/min. Data analysis was performed with Thermo Xcalibur (v3.0.63) and Qual Browser applications. Relative abundances of ethyl esters were determined by calculating the corresponding ethyl ester peak area over the total acyl chain peak area.

Phylogenetic analysis

Phylogenetic constructions were carried out using MEGA X (Kumar et al., 2018). Amino acid sequences were aligned with MUSCLE (Edgar, 2004) under default parameters. A maximum likelihood method based on Jones-Taylor-Thornton (JTT)+G model (Jones et al., 1992) was used to generate the phylogenetic trees, with 1000 bootstrap replicates. Phylogenetic trees were drawn to scale, with branch lengths computed using the Poisson correction method (Zuckerkandl and Pauling, 1965) in the number of amino acid substitutions per site.

CHAPTER V

CONCLUSIONS AND FUTURE DIRECTIONS

Acylsugars are nonvolatile and viscous plant specialized metabolites secreted through glandular trichomes on the surface of tomatoes and other plants of the nightshade family. These chemicals have documented insecticidal characteristics, as they deter herbivores and pests from damaging plants. In tomato breeding programs, manipulating acylsugar content and composition for better insect resistance has been a long-time goal. To make acylsugars, the plants attach carbon chains known as fatty acyl groups to sugar, such as sucrose. Some enzymes contributing to acylsugar synthesis were characterized in tomato, including a group of BAHD acyltransferases known as acylsucrose acyltransferases, and an invertase-like enzyme acylsucrose β-fructofuranosidase, while several acylsugar metabolic genes remain unidentified, and little is known about regulation of this pathway.

In this study, we exploited variation in acylsugar production among different *S. pennellii* accessions, and within a segregating F2 population from *S. pennellii* LA 0716 and *S. lycopersicum* cv. VF36 cross, through a combination of comparative transcriptomics analysis, biochemical inhibition study, and evolutionary analysis to identify a number of genes with known and putative functions in acylsugar metabolism. Additionally, we used virus-induced gene silencing methods to confirm two candidate genes, *SpKAR* and *SpRbcS*, to be involved in medium chain fatty acids synthesis pathway and carbon fixation associated with acylsugar production, respectively. These

results provide a strong support for the involvement of novel candidate genes we have identified in acylsugar biosynthesis, and serve as a valuable resource in future that could be used for engineering other crop plants to produce acylsugars as natural pesticides.

Future direction

1. Functional validation of more candidate genes involved in acylsugar production

We identified *Sopen05g009610, Sopen07g006810, Sopen05g032580* and *Sopen05g034770* in the intersection of three gene lists: DEGs in "HIGH"-and "LOW"-F2, DEGs in high- and low-acylsugar–producing *S. pennellii* accessions and genes with dN/dS>1 between *S. pennellii* and *S. lycopersicum* putative orthologs. We used VIGS to confirm the role of *Sopen05g009610* and *Sopen07g006810* in acylsugar production. For the gene *Sopen05g034770,* because its expression level was inversely correlated to acylsugar accumulation, and its low expression level in LA0716 made it a poor candidate for VIGS analysis, we did not study it further.

Sopen05g032580 was annotated as TUB8-like induced stolen tip protein, positively correlated to acylsugar accumulation, and rapidly evolving. This gene is also preferentially expressed in trichomes relative to the underlying tissue according to published trichome RNA sequencing data (Fan et al., 2020). I examined the sequence of *Sopen05g032580* and found conserved domains related to polynucleotide phosphorylase (PNPase) binding site. I am preparing VIGS targeting of this gene to determine if it is involved in acylsugar metabolism.

Meanwhile, we identified *Sopen05g008450* (Mandal et al., 2020), an AP2-family transcription factors, that was strongly connected with acylsugar related genes, such as three ASAT genes and the invertase gene, using weighted gene correlation network analysis (WGCNA; Langfelder and Horvath, 2008). We applied VIGS to this gene in *S. pennellii* LA0716, and the LC-MS results showed that total acylsugar levels decreased by 70% in *Sopen05g008450* VIGS group, compared to empty vector control group (unpublished data). We also used VIGS to test the in vivo role of *Sopen11g007710*, a peroxisomal membrane protein PMP22/Mpv17, and observed a 40% decrease of total acylsugar amount in VIGS group, compared to empty vector control group (unpublished data). We will use RNA-seq and Chip-seq to identify genes connected with *Sopen05g008450* and *Sopen11g007710* to uncover additional candidate genes that are involved in the production of acylsugar and regulation of acylsugar synthesis pathways.

2. Further exploring of genes under positive selection

Although we used comparative genomic approaches in two independent experiments to investigate the biosynthesis of acylsugars in *S. pennellii*, it is important to mention that not all the genes involved in acylsugar production would be in the differentially expressed genes sets. Nonsynonymous substitutions may also result in neofunctionalization of an enzyme, with no change in expression of the encoded gene. In this study, I identified 732 genes with dN/dS >1.0 in the genome-wide dN/dS calculation of *S. pennellii* and *S. lycopersicum* putative orthologs.

I hypothesized that genes related to the high acylsugar phenotype might be under positive selection, and these gene may have a dN/dS ratio more than 1.0 because high acylsugar levels provide *S. pennellii* advantages towards adaptation to arid environment and resistance to insect herbivores. Other than *Sopen05g009610* and *Sopen07g006810* that were functional validated in this dissertation, there might be other unknown rapidly evolving genes in this dN/dS >1 list, but not in either of the two DEG sets, that are involved in acylsugar production. I will examine this dN/dS >1.0 list further to identify more genes and evolutionary mechanisms underlying acylsugar metabolism in Solanaceae.

Finally, the set of high dN/dS genes may include some genes that influence fruit characteristics, because the fruits of *S. pennellii* and *S. lycopersicum* are so different. Unlike cultivated tomato fruits, fruits of *S. pennellii* are much smaller, mostly green and covered by acylsugars. Identification of such genes may be useful for agronomic applications in the future.

REFERENCES

Afendi, F. M., Okada, T., Yamazaki, M., Hirai-Morita, A., Nakamura, Y., Nakamura, K., et al. (2012). Knapsack family databases: integrated metabolite-plant species databases for multifaceted plant research. Plant Cell Physiol. 53:e1.

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

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.

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.

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.

Borak, B., Ort, D.R., and Burbaum, J.J. (2013) Energy and carbon accounting to compare bioenergy crops. Curr Opin Biotechnol. 24(3): 369-75

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.

Burke, B.A., Goldsby, G., and Mudd, J.B. (1987) Polar epicuticular lipids of *Lycopersicon pennellii*. Phytochemistry. 26: 2567-2571

Dembitsky, V.M. (2004) Chemistry and biodiversity of the biologically active natural glycosides. Chem Biodivers. 1(5): 673-781.

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.

Edgar, R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 5: 113.

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.

Fan, P., Leong, B.J. and Last, R.L. (2019). Tip of the trichome: evolution of acylsugar metabolic diversity in Solanaceae. Current Opinion in Plant Biology 49: 8–16

Fan, P., Wang, PP., Lou, YR., Leong, B.J., Moore, B.M., Schenck, C.A., Cao, P., Brandizzi, F.B., Shiu, SH. and Last, R.L. (2020). Evolution of a plant gene cluster in Solanaceae and emergence of metabolic diversity. eLife 9: e56717

Fan, P., Miller, AM., Schilmiller, A.L., Liu, X., Ofner, I., Jones, AD., 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 U S A. 113(2): E239-48.

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.

Fiehn, O. (2002) Metabolomics – the link between genotypes and phenotypes. Plant Molecular Biology 48: 155–171

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.

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.

Hawkins, A.K., Garza, E.R., Dietz, V.A., Hernandez, O.J., Hawkins, W.D., Burrell, A.M., and Pepper, A.E. (2017) Transcriptome Signatures of Selection, Drift, Introgression, and Gene Duplication in the Evolution of an Extremophile Endemic Plant. Genome Biol Evol 9:3478-3494.

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

Jones, D.T., Taylor, W.R., and Thornton, J.M. (1992) The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. 8: 275–282.

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

Kandra, L., and Wagner, G.J. (1988) Studies of the site and mode of biosynthesis of tobacco trichome exudate components. Arch Biochem Biophys. 265: 425-32.

Kandra, L., and Wagner, G.J. (1990) Chlorsulfuron modifies biosynthesis of acyl Acid substituents of sucrose esters secreted by tobacco trichomes. Plant Physiol. 94(3): 906-12.

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.

King, R.R., Calhoun, L.A., and Singh, R.P. (1988) 3,4-Di-O- and 2,3,4-tri-*O*-acylated glucose esters from the glandular trichomes of nontuberous Solanum species. Phytochemistry. 27: 3765-3768

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(6): 1013-21.

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.

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.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol. 35(6): 1547-1549.

Lahti, D.C., Johnson, N.A., Ajie, B.C., Otto, S.P., Hendry, A.P., Blumstein, D.T., Coss, R.G., Donohue, K., and Foster, S.A. (2009). Relaxed selection in the wild. Trends Ecol Evol. 24: 487-496.

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

Laterre, R., Pottier, M., Remacle, C., and Boutry, M. (2017) Photosynthetic Trichomes Contain a Specific Rubisco with a Modified pH-Dependent Activity. Plant Physiol. 173(4): 2110-2120.

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.

Liu, Y., Jing, S.X., Luo, S.H., and Li, S.H. (2019) Non-volatile natural products in plant glandular trichomes: chemistry, biological activities and biosynthesis. Nat Prod Rep. 36(4): 626-665

Maeda, K., Kobayashi, Y., Koide, M., Uehara, S., Okamoto, M., Ishihara, A., Kayama, T., Saito, M., and Marumo, K. (2019) The Regulation of Bone Metabolism and Disorders by Wnt Signaling. Int J Mol Sci. 20(22): 5525.

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.

Mandal, S., Ji, W., and McKnight, T.D. (2020) Candidate Gene Networks for Acylsugar Metabolism and Plant Defense in Wild Tomato *Solanum pennellii*. Plant Cell. 32(1): 81-99.

Martin-Avila, E., Lim, Y.L., Birch, R., Dirk, L.M.A., Buck, S., Rhodes, T., Sharwood, R.E., Kapralov, M.V., and Whitney, S.M. (2020) Modifying Plant Photosynthesis and Growth via Simultaneous Chloroplast Transformation of Rubisco Large and Small Subunits. Plant Cell. 32(9): 2898-2916.

Martin, B., and Thorstenson, Y.R. (1988) Stable carbon isotope composition (deltaC), water use efficiency, and biomass productivity of *Lycopersicon esculentum*, *Lycopersicon pennellii*, and the F1 hybrid. Plant Physiol. 88: 213-7.

Milo, R., and Last, R.L. (2012) Achieving diversity in the face of constraints: lessons from metabolism. Science. 336(6089): 1663-7.

Mithöfer, A., and Boland, W. (2012) Plant defense against herbivores: chemical aspects. Annu Rev Plant Biol. 63: 431-50.

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(3): 1512-23.

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.

Montoya-Burgos, JI. (2011) Patterns of positive selection and neutral evolution in the protein-coding genes of Tetraodon and Takifugu. PLoS One. 6(9): e24800.

Mugal, C.F., Wolf, J.B., and Kaj, I. (2014) Why time matters: codon evolution and the temporal dynamics of dN/dS. Mol Biol Evol. 31: 212-231.

Mutschler, M.A., and Wintermantel, W.M. (2006). Reducing virus associated crop loss through resistance to insect vectors. Natural Resistance Mechanisms of Plants to Viruses. G. Loebenstein and J.P. Carr (eds.)241-260.

Ndimba, B.K., Ndimba, R.J., Johnson, T.S., Waditee-Sirisattha, R., Baba, M., Sirisattha, S., Shiraiwa, Y., Agrawal, G.K., and Rakwal, R. (2013) Biofuels as a sustainable energy source: an update of the applications of proteomics in bioenergy crops and algae. J Proteomics. 93: 234-44.

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.

Panchy, N., Lehti-Shiu, M., and Shiu, SH. (2016) Evolution of Gene Duplication in Plants. Plant Physiol. 171(4): 2294-316.

Pichersky, E., and Lewinsohn, E. (2011) Convergent evolution in plant specialized metabolism. Annu Rev Plant Biol. 62: 549-66.

Pottier, M., Gilis, D., and Boutry, M. (2018) The Hidden Face of Rubisco. Trends Plant Sci. 23(5): 382-392.

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.

Pyke, K.A., and Howells, C.A. (2002) Plastid and stromule morphogenesis in tomato. Ann Bot. 90(5): 559-66.

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.

Savchenko, T., Whiting, P., Germade, A., and Dinan, L. (2000) Ecdysteroid agonist and antagonist activities in species of the Solanaceae. Biochem Syst Ecol. 28(5): 403-419.

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.

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.

Shepherd, R.W., Bass, W.T., Houtz, R.L., and Wagner, G.J. (2005) Phylloplanins of tobacco are defensive proteins deployed on aerial surfaces by short glandular trichomes. 17(6): 1851-61.

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.

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.

Strasburg, J.L, Kane, N.C., Raduski, A.R., Bonin, A., Michelmore, R., and Rieseberg, L.H. (2011) Effective population size is positively correlated with levels of adaptive divergence among annual sunflowers. Mol Biol Evol. 28: 1569-1580.

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.

van der Hoeven, R.S., and Steffens, J.C. (2000) Biosynthesis and elongation of short- and medium-chain-length fatty acids. Plant Physiol. 122(1): 275-82.

Wagner, G.J., Wang, E., and Shepherd, R.W. (2004) New approaches for studying and exploiting an old protuberance, the plant trichome. Ann Bot. 93: 3–11

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.

Warnock, S.J. (1991) Natural habitats of Lycopersicon species. Hortscience. 26: 466-71.

Wu, Q., Cho, J.G., Lee, D.S., Lee, D.Y., Song, N.Y., Kim, Y.C., Lee, K.T., Chung, H.G., Choi, M.S., Jeong, T.S., et al. (2013) Carbohydrate derivatives from the roots of *Brassica rapa ssp. campestris* and their effects on ROS production and glutamate-induced cell death in HT-22 cells. Carbohydr Res. 372: 9-14.

Yang, Z. (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. Computer applications in the biosciences. CABIOS 13: 555-556.

Yang, Z., and Bielawski, J.P. (2000) Statistical methods for detecting molecular adaptation. Trends in Ecology & Evolution. 15 (12): 496–503.

Yang, Z., and Nielsen, R. (2000) Estimating Synonymous and Nonsynonymous Substitution Rates Under Realistic Evolutionary Models. Molecular Biology and Evolution. 17: 32-43.

Yu, A.T. (1972) The genetics and physiology of water usage in *Solanum pennellii* Corr. and its hybrids with *Lycopersicon esculentum* Mill. PhD thesis. University of California, Davis.

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.

Zuckerkandl, E., and Pauling, L. (1965) Molecules as documents of evolutionary history. J Theor Biol. 8(2): 357-66.

APPENDIX A

All of the Supplemental Datasets can be found from the link below:

https://github.com/feelingice/Dissertation-Supplemental-Datasets