https://doi.org/10.1093/plphys/kiac487 PLANT PHYSIOLOGY 2023: 191: 1199–1213

Plant Physiology®

Research Article

# Specialized metabolism by trichome-enriched Rubisco and fatty acid synthase components

Wangming Ji [1], Sabyasachi Mandal [1], Yohannes H. Rezenom [2] and Thomas D. McKnight [1,*,†]

1 Department of Biology, Texas A&M University, College Station, Texas 77843, USA
2 Department of Chemistry, Texas A&M University, College Station, Texas 77843, USA

*Author for correspondence: mcknight@bio.tamu.edu
†Senior author

These authors contributed equally (W.J. and S.M.).

W.J. and S.M. designed and performed experiments, analyzed data, and wrote the manuscript. Y.H.R. designed and performed chromatography–mass spectrometry analyses, and reviewed the manuscript. T.D.M. designed experiments, analyzed data, and wrote the manuscript. All authors read and approved the final manuscript.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (https://academic.oup.com/plphys/pages/general-instructions) is Thomas D. McKnight (mcknight@bio.tamu.edu).

## Abstract

Acylsugars, specialized metabolites with defense activities, are secreted by trichomes of many solanaceous plants. Several acylsugar metabolic genes (AMGs) remain unknown. We previously reported multiple candidate AMGs. Here, using multiple approaches, we characterized additional AMGs. First, we identified differentially expressed genes between high- and low-acylsugar-producing $F_2$ plants derived from a cross between cultivated tomato (*Solanum lycopersicum*) and a wild relative (*Solanum pennellii*), which produce acylsugars that are ∼1% and ∼20% of leaf dry weight, respectively. Expression levels of many known and candidate AMGs positively correlated with acylsugar amounts in $F_2$ individuals. Next, we identified *lycopersicum-pennellii* putative orthologs with higher nonsynonymous to synonymous substitutions. These analyses identified four candidate genes, three of which showed enriched expression in stem trichomes compared to underlying tissues (shaved stems). Virus-induced gene silencing confirmed two candidates, *Sopen05g009610* [beta-ketoacyl-(acyl-carrier-protein) reductase; fatty acid synthase component] and *Sopen07g006810* (Rubisco small subunit), as AMGs. Phylogenetic analysis indicated that *Sopen05g009610* is distinct from specialized metabolic cytosolic reductases but closely related to two capsaicinoid biosynthetic reductases, suggesting evolutionary relationship between acylsugar and capsaicinoid biosynthesis. Analysis of publicly available datasets revealed enriched expression of *Sopen05g009610* orthologs in trichomes of several acylsugar-producing species. Similarly, orthologs of *Sopen07g006810* were identified as solanaceous trichome-enriched members, which form a phylogenetic clade distinct from those of mesophyll-expressed "regular" Rubisco small subunits. Furthermore, $\delta^{13}C$ analyses indicated recycling of metabolic $CO_2$ into acylsugars by Sopen07g006810 and showed how trichomes support high levels of specialized metabolite production. These findings have implications for genetic manipulation of trichome-specialized metabolism in solanaceous crops.

## Introduction

Plant metabolites are traditionally classified into primary or central metabolites and secondary or specialized metabolites. In contrast to evolutionarily conserved primary metabolites, specialized metabolites are found in specific taxonomic groups and exhibit greater structural diversity. The building blocks of specialized metabolites are derived from products of primary metabolism; for example, alkaloids are derived from

Received July 12, 2022. Accepted September 26, 2022. Advance access publication October 20, 2022


Open Access

amino acids, whereas acylsugars are derived from sugar and fatty acids (Walters and Steffens, 1990; Fan et al., 2016). The evolution of a specialized metabolic pathway requires evolution of new gene functions, which can be achieved through a variety of mechanisms, including duplication of primary metabolic genes followed by neo- or sub-functionalization and/or changes in spatio-temporal gene expression (Moghe and Last, 2015).

Specialized metabolites have important roles in plant–environment interactions. For example, acylsugars, which are nonvolatile and viscous metabolites secreted through glandular trichomes of many species in the Solanaceae (Slocombe et al., 2008; Moghe et al., 2017), provide protection against biotic and abiotic stress. These compounds contribute directly and indirectly to plant defense by providing resistance against insect herbivores (Alba et al., 2009; Leckie et al., 2016), by mediating multitrophic defense by attracting predators of herbivores through volatile short-chain aliphatic acids produced from acylsugar breakdown (Weinhold and Baldwin, 2011), and by protecting plants from microbial pathogens (Luu et al., 2017). Acylsugars also protect plants from desiccation (Fobes et al., 1985; Feng et al., 2021). These beneficial properties have led to interest in understanding acylsugar metabolism and identifying factors that control their production for breeding agronomically important crops with better resistance against insect herbivores (Bonierbale et al., 1994; Lawson et al., 1997; Leckie et al., 2012).

Acylsugars exhibit tremendous structural variation in the Solanaceae. Both branched- and straight-chain fatty acids (SCFAs) are esterified to the sugar moiety (glucose or sucrose) to form acylsugars, and major acyl chains vary in length (C2–C12) in different species (Kroumova et al., 2016; Moghe et al., 2017). Predominant branched-chain fatty acids include 2-methylpropanoate, 3-methylbutanoate, and 2-methylbutanoate, which are derived from branched-chain amino acids (valine, leucine, and isoleucine, respectively) (Walters and Steffens, 1990). Branched medium-chain acyl groups, such as 6-methylheptanoate and 8-methylnonanoate, are derived from branched short-chain precursors through elongation reactions mediated by either α-ketoacid (one-carbon elongation) or fatty acid synthase (FAS; two-carbon elongation) (Kroumova and Wagner, 2003). Predominant SCFAs, such as *n*-decanoate and *n*-dodecanoate, are presumably derived from acetyl-CoA via FAS-mediated de novo biosynthesis (Walters and Steffens, 1990; Mandal et al., 2020). Once acyl chains are produced, specific sets of acylsugar acyltransferases (ASATs) in different species of the Solanaceae attach these aliphatic groups to different carbon positions of the sugar moiety, leading to remarkable metabolic diversity (Schilmiller et al., 2015; Fan et al., 2016; Moghe et al., 2017; Nadakuduti et al., 2017; Feng et al., 2021; Chang et al., 2022; Schenck et al., 2022). Although many acylsugar metabolic genes (AMGs) have been identified in recent years, several remain unidentified and the regulation of acylsugar biosynthesis has not been well characterized, especially with regard to how solanaceous trichomes can support production of high levels of specialized metabolites.

*Solanum pennellii*, a wild relative of the cultivated tomato (*Solanum lycopersicum*), is endemic to arid western slopes of the Peruvian Andes and is known for its ability to withstand extreme drought conditions. Acylsugars secreted from glandular trichomes of *S. pennellii* represent a remarkably large fraction, up to 20%, of the leaf dry weight; in contrast, *S. lycopersicum* produces only ∼1% of its leaf dry weight as acylsugars (Fobes et al., 1985). Here, in order to identify AMGs, we first created a *S. lycopersicum* × *S. pennellii* $F_2$ population and conducted transcriptomic comparisons between high- and low-acylsugar-producing $F_2$ individuals to identify candidate AMGs. Next, we compared this list with our previously published list of candidate AMGs (Mandal et al., 2020). Additionally, genome-wide nonsynonymous to synonymous substitution rates (dN/dS ratio) estimation of *S. pennellii* and *S. lycopersicum* putative orthologs refined the list of candidate genes, and analysis of trichome-enriched expression identified three candidates: *Sopen05g009610*, encoding a beta-ketoacyl-(acyl-carrier-protein) reductase (SpKAR1 hereafter; a component of the FAS complex), *Sopen07g006810*, encoding a small subunit of Rubisco (SpRBCS1 hereafter), and *Sopen05g032580*, encoding an induced stolon-tip protein like member (SpSTPL hereafter). Virus-induced gene silencing (VIGS) indicated roles for both *SpKAR1* and *SpRBCS1* in acylsugar metabolism, while silencing of *SpSTPL* had no effect. The role of SpRBCS1 was further supported by $\delta^{13}C$ analyses of acylsugars. Additionally, phylogenetic analyses and investigation of publicly available large-scale datasets revealed interesting evolutionary aspects which suggested that orthologs of SpKAR1 and SpRBCS1 are involved in specialized metabolic pathways in other plants.

## Results

### Transcriptomic comparison between high- and low-acylsugar-producing $F_2$ individuals of *S. lycopersicum* × *S. pennellii*

*Solanum pennellii* accession LA0716 produces copious amount of acylsugars (∼20% of leaf dry weight), whereas *S. lycopersicum* cv. VF36 accumulates considerably lower amount (∼1% of leaf dry weight) (Fobes et al., 1985). The interspecific $F_1$ hybrid (LA4135) accumulates low levels of acylsugars (<3% of leaf dry weight). To identify candidate AMGs, we first analyzed acylsugar accumulation in an $F_2$ population derived from the cross between VF36 and LA0716. Of the 114 $F_2$ plants, 24 accumulated acylsugars >10% of their leaf dry weight, whereas 62 accumulated acylsugars <3% of their leaf dry weight (Figure 1A). Next, using RNA-sequencing (RNA-seq), we identified genes that were differentially expressed between 10 high- and 10 low-acylsugar-producing $F_2$ individuals (hereafter referred to as HIGH-$F_2$ and LOW-$F_2$, respectively). A total of 20,160 *S. pennellii* genes were selected after minimum-expression-level filtering, and 331 differentially expressed genes (DEGs) were identified; of these, 134 and 197

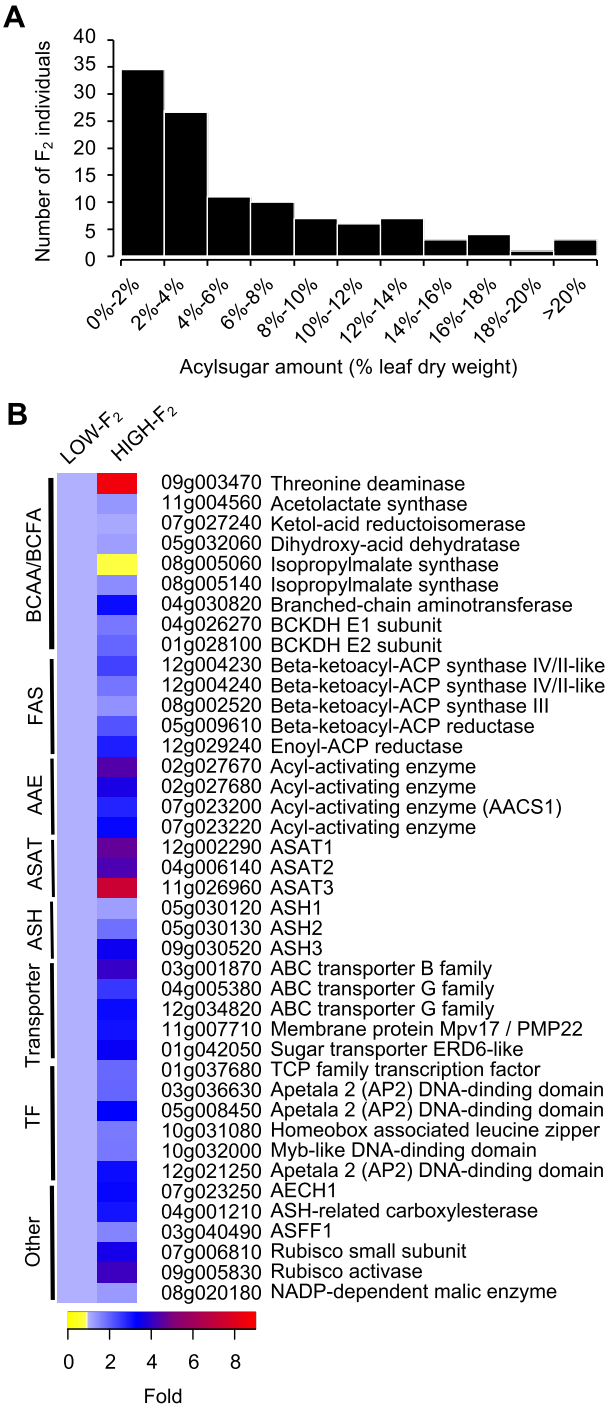


**Figure 1** Acylsugar accumulation and expression of known and candidate AMGs in a *S. lycopersicum* VF36 × *S. pennellii* LA0716 $F_2$ population. A, Histogram showing acylsugar amount distribution among 114 $F_2$ plants. For each plant, three replicates were used to measure acylsugar amount. B, Heatmap showing relative expression levels (set to one-fold in the LOW-$F_2$ group) of genes with known and putative roles in acylsugar metabolism. *Solanum pennellii* gene identifier numbers (Sopen IDs) are given with annotations. BCAA, branched-chain amino acid; BCFA, branched-chain fatty acid; FAS, fatty acid synthase component; AAE, acyl-activating enzyme; ASAT, acylsugar acyltransferase; ASH, acylsugar acylhydrolase; TF, transcription factor; BCKDH, branched-chain keto acid dehydrogenase; AACS1, acylsugar acyl-CoA synthetase 1; AECH1, acylsugar enoyl-CoA hydratase 1; ASFF1, acylsucrose fructofuranosidase 1.

DEGs showed higher and lower expression levels, respectively, in the HIGH-$F_2$ group compared to the LOW-$F_2$ group (Supplemental Data Set 1). Enrichment analysis indicated that gene ontology (GO) terms such as "acyltransferase activity" (GO:0016747), "fatty acid metabolic process" (GO:0006631), and "active transmembrane transporter activity" (GO:0022804) were over-represented in the list of 331 DEGs (Supplemental Figure S1).

Of the 331 DEGs, 73 were also differentially expressed between high- and low-acylsugar-producing accessions of *S. pennellii* (Mandal et al., 2020). Many genes with known and putative roles in acylsugar metabolism showed higher expression levels in the HIGH-$F_2$ group compared to the LOW-$F_2$ group (Figure 1B), which validated our RNA-seq approach. These genes encode branched-chain fatty acid metabolic proteins, FAS components, acyl-activating enzymes (also known as acyl-CoA synthetases), a mitochondrial/peroxisomal membrane protein (Sopen11g007710), three ATP-binding cassette transporters, three acylsugar acyltransferases (ASATs), transcription factors, and a Rubisco small subunit. Additionally, known and putative flavonoid metabolic genes, such as *Sopen11g003320* (UDP-glucose:catechin glucosyltransferase) and three sequential genes on chromosome 6 (*Sopen06g034810*, *Sopen06g034820*, and *Sopen06g034830*; myricetin *O*-methyltransferase; Kim et al., 2014), which are strongly co-expressed with AMGs (Mandal et al., 2020), were also found in the list of DEGs (Supplemental Data Set 1).

## Genome-wide dN/dS ratio estimation of *S. pennellii* and *S. lycopersicum* putative orthologs and refinement of candidate AMG list

In addition to difference in gene expression levels, difference in protein-coding sequence may also contribute to difference in acylsugar accumulation capabilities between *S. pennellii* and *S. lycopersicum*. Compared to primary metabolic genes, specialized metabolic genes evolve faster and exhibit higher ratios of nonsynonymous to synonymous substitution rates (dN/dS) (Moore et al., 2019). Therefore, as an additional approach for identifying candidate AMGs, we performed a genome-wide dN/dS ratio analysis (Yang and Nielsen, 2000) of *S. pennellii* and *S. lycopersicum* putative orthologs to identify genes that are under positive selection. Using reciprocal BLAST, 19,984 putative ortholog pairs were selected (Supplemental Data Set 2), and the yn00 maximum-likelihood method (Yang, 1997) yielded a genome-wide mean dN/dS ratio of 0.3273 (Figure 2A). A total of 732 genes with dN/dS $>$ 1.0 were considered to be under positive selection (Supplemental Data Set 3).

To narrow our focus further, we next compared lists of candidate AMGs obtained from three approaches: (1) 331 DEGs identified in our analysis of the $F_2$ population; (2) 1,087 DEGs, we previously reported from transcriptomic analysis between high- and low-acylsugar-producing *S. pennellii* accessions (Mandal et al., 2020); and (3) 732 genes with dN/dS $>$ 1.0. Four genes (*Sopen07g006810*, *Sopen05g009610*, *Sopen05g032580*, and *Sopen05g034770*) occurred in all three sets (Figure 2B; Table 1). Next, we measured trichome-enriched expression of these four candidates, since many AMGs are expressed in trichome tip-cells (Ning et al., 2015; Schilmiller et al., 2015; Fan et al., 2016, 2020). Transcript levels of *Sopen07g006810* (*SpRBCS1*), *Sopen05g009610* (*SpKAR1*), and *Sopen05g032580* (*SpSTPL*) were 220-, 110-, and 13-fold higher, respectively, in isolated stem trichomes than in underlying tissues of shaved stems in *S. pennellii* accession LA0716. On the other hand, expression of *Sopen05g034770* was lower in trichomes (Figure 2C). Because of this, and the fact that expression of *Sopen05g034770* is inversely correlated with acylsugar amount (Table 1), this gene was not analyzed further.

## In vivo functional validation of candidate AMGs

To determine whether *SpRBCS1*, *SpKAR1*, and *SpSTPL* are involved in acylsugar biosynthesis, these three candidate genes were targeted in *S. pennellii* LA0716 for VIGS using tobacco rattle virus (TRV)-based silencing vectors (Dong et al., 2007). VIGS resulted in significant downregulation of target genes (82%, 54%, and 94% reduction in transcript levels for *SpRBCS1*, *SpKAR1*, and *SpSTPL*, respectively; Supplemental Figure S2, A and B), and leaf surface metabolites were analyzed by liquid chromatography–mass spectrometry (LC–MS). Compared to a group of control plants (empty TRV vectors), total acylsugar levels decreased by 23% ($P < 0.05$) in VIGS-*SpRBCS1* plants (Figure 3, A and B; Supplemental Data Set 4). In contrast, we did not observe any statistically significant changes in total acylsugar amount upon suppression of *SpKAR1* or *SpSTPL* (Figure 3A). Furthermore, suppression of *SpRBCS1* also led to a decrease in the amount of a nonacylsugar compound that appears to be a flavonoid, but suppression of *SpKAR1* or *SpSTPL* had no effect on the amount of this compound (Figure 3C).

Despite the lack of influence on total acylsugar amount, silencing of *SpKAR1* led to significant reductions in acylsugar medium-chain fatty acids (18%, 23%, and 20% reductions for 8-methylnonanoate, *n*-decanoate, and *n*-dodecanoate, respectively), which is consistent with its predicted role in medium-chain fatty acid biosynthesis (Figure 3D). Silencing of *SpRBCS1* resulted in 20% reductions in C5 acyl chains (2- and 3-methylbutanoate). Morphological differences were not observed between control and silenced plants (Supplemental Figure S2C), which indicates that these effects on acylsugar phenotypes were not indirect consequences of abnormal plant growth and development caused by VIGS. Transcript enrichment in trichomes and VIGS results together indicate roles of SpRBCS1 and SpKAR1 in acylsugar metabolism. On the other hand, no statistically significant changes in acyl chain profile were observed upon silencing of *SpSTPL*.

To further investigate the effects of *SpRBCS1* and *SpKAR1* silencing on individual acylsugars, we quantified these peaks using LC–MS. While several acylsugar peaks showed reductions in VIGS-*SpRBCS1* plants, no significant reductions were observed in VIGS-*SpKAR1* plants (Supplemental Figure S2D). This is likely due to the fact that at least two of the three acyl groups in *S. pennellii* acylsugars are short chain, which are not targets of SpKAR1.

## Phylogenetic analysis of SpKAR1

Short-chain dehydrogenases/reductases (SDRs) constitute a large protein superfamily of NAD(P)(H)-dependent oxidoreductases; members exhibit low levels of sequence identity,

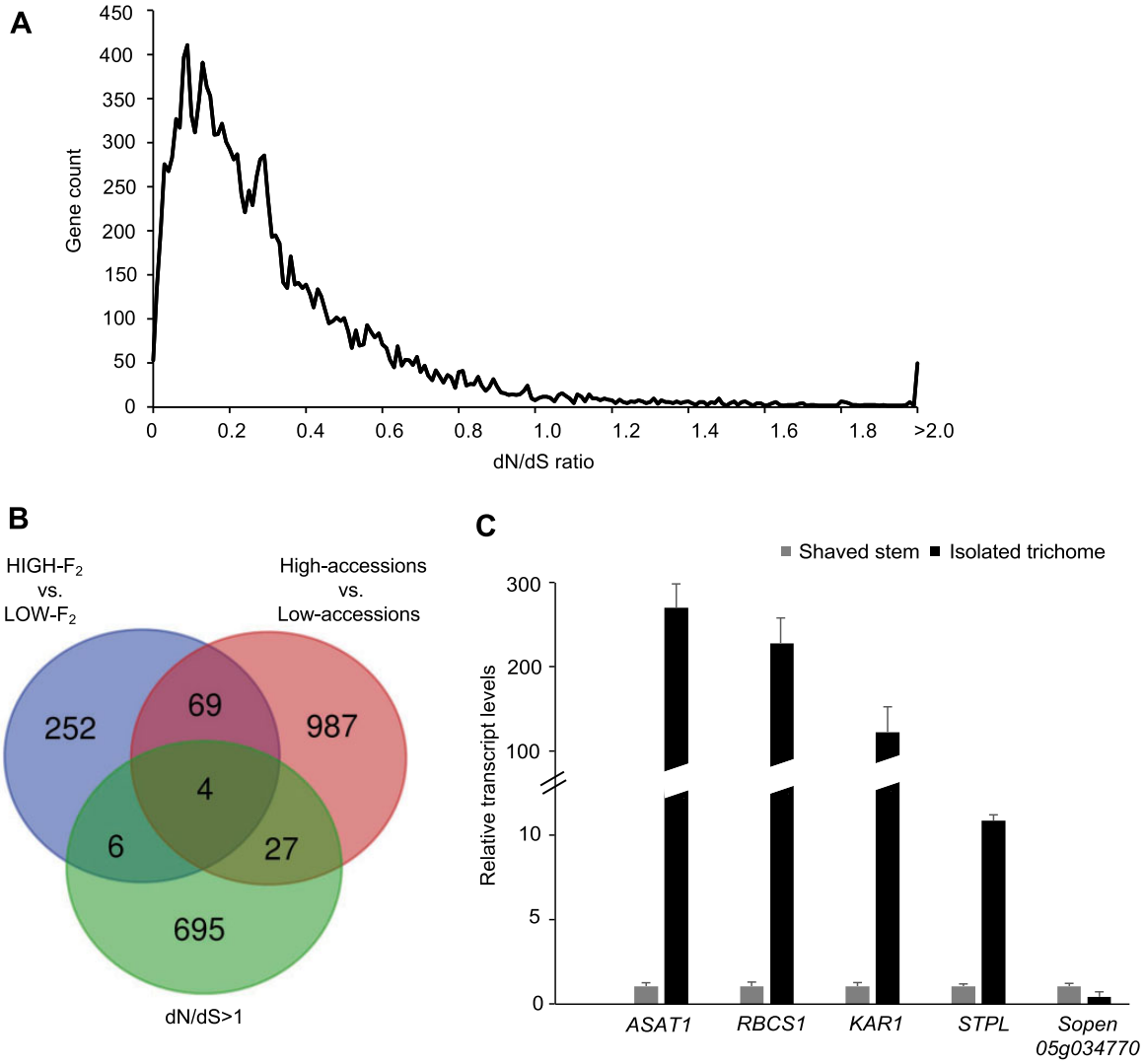


**Figure 2** Selection of candidate AMGs. A, Distribution of nonsynonymous to synonymous substitution rate ratios (dN/dS ratios) of putative ortholog pairs from *S. pennellii* and *S. lycopersicum*. B, Venn diagram showing the intersections of three sets of candidate AMGs: (1) DEGs between high- and low-acylsugar-producing *S. pennellii* accessions (Mandal et al., 2020); (2) DEGs between high- and low-acylsugar-producing $F_2$ plants (HIGH-$F_2$ and LOW-$F_2$, respectively); and (3) genes with dN/dS > 1 between *S. pennellii* and *S. lycopersicum* putative orthologs. Four genes were identified at the intersections of these three sets. C, Relative expression levels of the four candidate AMGs [*SpRBCS1* (*Sopen07g006810*), *SpKAR1* (*Sopen05g009610*), *SpSTPL* (*Sopen05g032580*), and *Sopen05g034770*] in isolated stem trichomes and underlying tissues of shaved stems (normalized to one-fold) in *S. pennellii* LA0716. *SpASAT1* (*Sopen12g002290*), the ortholog of *S. lycopersicum* trichome tip-cell-expressed *ASAT1* (Fan et al., 2016), was included for comparison. Error bars indicate SE ($n$ = 5 individual plants).

**Table 1** Four genes identified at the intersections of three sets of candidate AMGs

| Gene ID | High versus Low Accessions $log_2$FC | HIGH-$F_2$ versus LOW-$F_2$ $log_2$FC | dN/dS Value | dN/dS Rank | Annotation |
|---|---|---|---|---|---|
| Sopen07g006810 | 5.01 | 1.85 | 2.91 | 29 | Rubisco small subunit |
| Sopen05g009610 | 3.32 | 1.15 | 2.8 | 35 | Beta-ketoacyl-(acyl-carrier-protein) reductase |
| Sopen05g032580 | 2.97 | 2.1 | 1.52 | 227 | Induced stolon tip protein-like |
| Sopen05g034770 | −3.48 | −1.89 | 1.59 | 199 | Uncharacterized protein |

$Log_2$FC indicates $log_2$ (fold-change). Positive and negative $log_2$FC values indicate higher and lower expression levels, respectively, in high-acylsugar-producing *S. pennellii* accessions (Mandal et al., 2020) or HIGH-$F_2$ group. dN/dS rank represents the corresponding rank of selected gene in 732 putative orthologs with dN/dS > 1.

but they share a Rossmann-fold motif for nucleotide binding (Kavanagh et al., 2008). Many SDRs are involved in biosynthesis of specialized metabolites, such as sesquiterpene zerumbone in bitter ginger (*Zingiber zerumbet*) (Okamoto et al., 2011), diterpene momilactone in rice (*Oryza sativa*) (Shimura et al., 2007), monoterpenoids in glandular

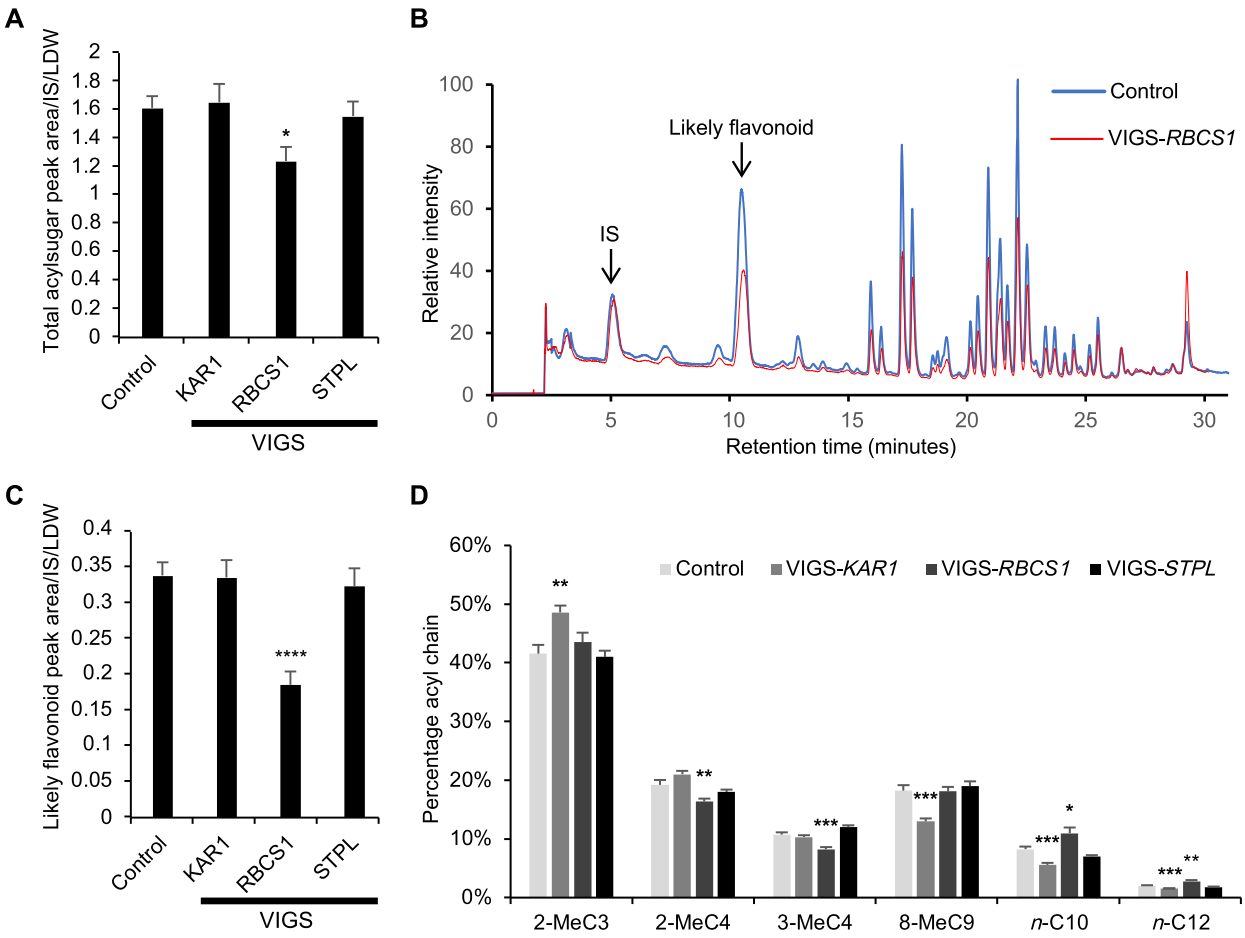


**Figure 3** VIGS of three trichome-enriched candidate AMGs in *S. pennellii* LA0716. For (A), (C), and (D), error bars indicate SE ($n$ = 10, 12, 11, and 12 individual plants for control, VIGS-*KAR1*, VIGS-*RBCS1*, and VIGS-*STPL* groups, respectively; $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$; Dunnett's test). A, Acylsugar quantification by LC–MS. To quantify acylsugar amounts, chromatogram peak areas were normalized by internal standard (IS) area and leaf dry weight (LDW). B, Representative chromatograms (normalized by IS area and LDW) showing metabolite peaks in control and VIGS-*RBCS1* plants. Metabolite peaks are listed in Supplemental Data Set 4. C, Quantification of a likely flavonoid peak in LC–MS analysis. D, Acylsugar acyl chain composition analysis by GC–MS. Predominant acyl chains are shown. Me, methyl; C3-C12 indicate acyl chain length (e.g. 2-MeC3 and *n*-C10 indicate 2-methylpropanoate and *n*-decanoate, respectively).

trichomes of sweet wormwood (*Artemisia annua*) (Polichuk et al., 2010), monoterpenoid constituents of essential oils in peppermint (*Mentha* × *piperita*) and spearmint (*Mentha spicata*) (Ringer et al., 2005), phenolic monoterpenes in the Lamiaceae (Krause et al., 2021), and steroidal glycoalkaloids and saponins in *Solanum* species (Sonawane et al., 2018). SpKAR1 (322 aa) belongs to the SDR family and shares several common motifs, such as the N-terminal cofactor-binding motif TGxxxGxG, a downstream structural motif (C/N)NAG, the active site motif YxxxK, and the catalytic tetrad N-S-Y-K, with members of the SDR "classical" subfamily (Kavanagh et al., 2008) (Supplemental Figure S3). However, phylogenetic analysis suggested that SpKAR1 is closer to *Escherichia coli* and *Synechocystis* sp. PCC 6803 FabG [beta-ketoacyl-(acyl-carrier-protein) reductase] than to other specialized metabolic SDRs mentioned earlier (Figure 4A; Supplemental Figure S4). Additionally, these plant SDRs are predicted to have a cytosolic location, whereas SpKAR1 is predicted to be localized to the chloroplast [TargetP (https://services.healthtech.dtu.dk/service.php?TargetP-2.0), WoLF PSORT (https://www.genscript.com/wolf-psort.html), and MultiLoc2 (https://abi-services.informatik.uni-tuebingen.de/multiloc2/webloc.cgi)]. Furthermore, cytochrome P450 monooxygenase (CYP) activities are closely associated with these specialized metabolic SDRs, whereas SpKAR1 activity presumably is not associated with CYP products. These results together indicate that SpKAR1 is a component of the FAS complex in the chloroplast, and it is phylogenetically distant from other specialized metabolic SDRs mentioned earlier.

Bell pepper (*Capsicum annuum*) is a solanaceous species that produces capsaicinoid specialized metabolites (Mazourek et al., 2009), but no detectable acylsugars were reported in this species (Moghe et al., 2017). Two SDRs from *C. annuum* showed close phylogenetic relationships with SpKAR1, and we investigated the similarity between capsaicinoid and acylsugar biosynthetic pathways. In both pathways, valine is converted to 2-methylpropanoyl-CoA,

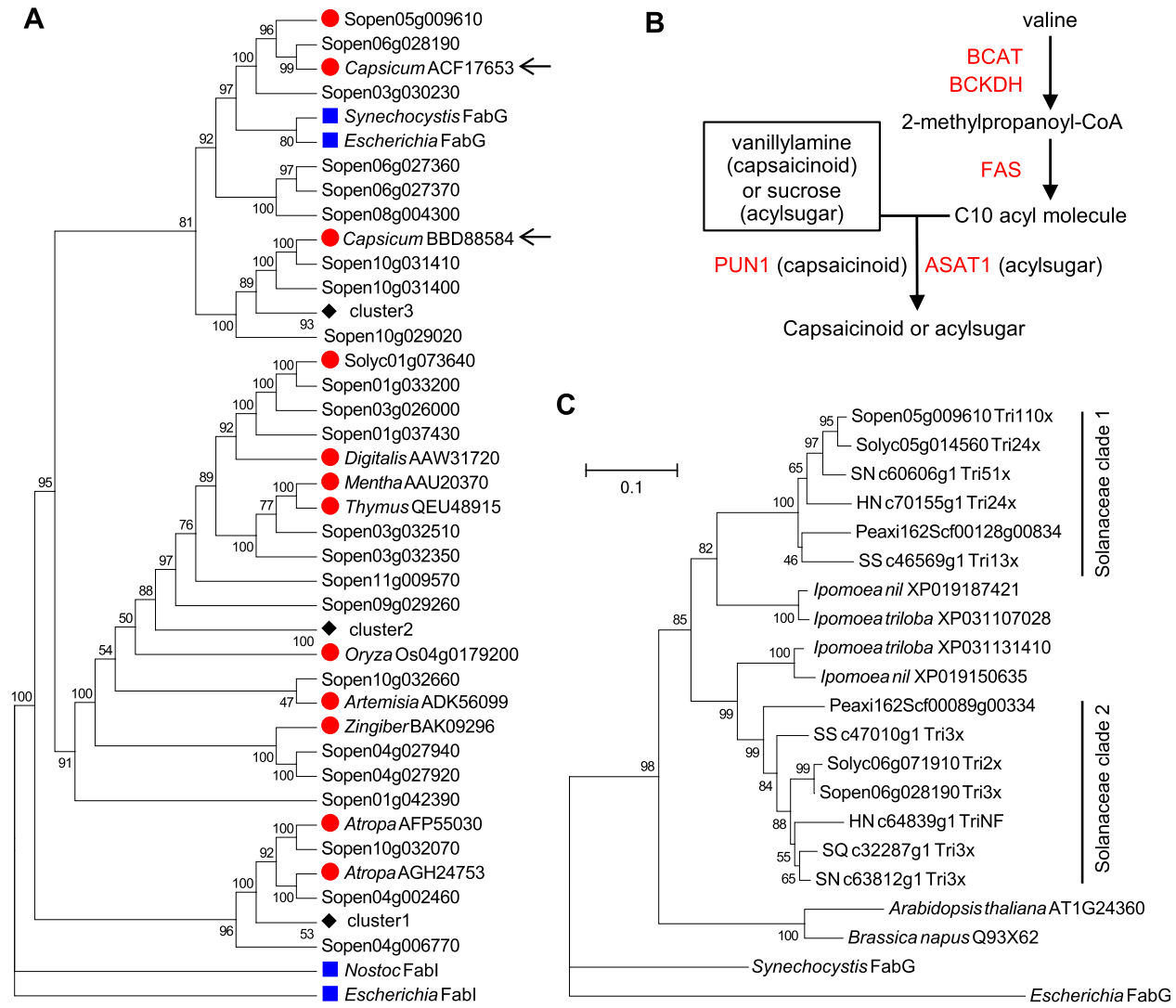


**Figure 4** Phylogenetic analyses of SpKAR1. A, Maximum-likelihood tree (topology) of SpKAR1 (Sopen05g009610; highlighted) and related SDRs. Red circles indicate plant-specialized metabolic SDRs. Blue squares indicate bacterial sequences. "Sopen" numbers indicate sequences from *S. pennellii*. Sequences from other species are given with GenBank accession numbers. Black diamonds indicate more than one "Sopen" sequences, which were clustered to save space; complete tree is given in Supplemental Figure S4. Two sequences related to capsaicinoid biosynthesis are indicated by arrows. Bootstrap values from 1,000 replicates are shown on the nodes. B, Similarity between capsaicinoid and acylsugar biosynthetic pathways. Metabolism of leucine, isoleucine, and SCFAs in acylsugar pathway are not shown here. Single arrows do not necessarily indicate single enzymatic steps. Enzymes are in red font. BCAT, branched-chain aminotransferase; BCKDH, branched-chain keto acid dehydrogenase; FAS, fatty acid synthase; PUN1, pungent gene 1; ASAT1, acylsugar acyltransferase 1. C, Neighbor-joining tree of SpKAR1 (highlighted) and its homologs in the Solanaceae. Sequences from four nonsolanaceous plant species [*Ipomoea triloba* and *I. nil* (Convolvulaceae); *A. thaliana* and *Brassica napus* (Brassicaceae)] and two bacteria (*Synechocystis* and *Escherichia*) are also included. Bootstrap values from 1,000 replicates are shown. Tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Tri110x indicates 110-fold higher expression in isolated trichomes compared to underlying tissues (NF, not found in HN_c64839g1). RT–qPCR was used for "Sopen" sequences. Trichome-enriched expression data (based on RNA-seq) for sequences in five other species were obtained from Ning et al. (2015) (Solyc, *S. lycopersicum*) and Moghe et al. (2017) (SN, *S. nigrum*; SQ, *S. quitoense*; HN. *H. niger*; SS, *S. sinuata*). Peaxi. *P. axillaris*. Sopen03g030230 (138 aa) and its putative orthologs were not included because they have long deletions and insertions. Maximum-likelihood tree is given in Supplemental Figure S5.

which is then elongated to C10 acyl molecules via FAS-mediated reactions (Figure 4B). This suggests a common evolutionary origin for acyl chain elongation steps in both acylsugar and capsaicinoid biosynthesis.

Ning et al. (2015) reported trichome-enriched expression data for *S. lycopersicum* genes, and Moghe et al. (2017) reported similar data in four additional acylsugar-producing solanaceous species: black nightshade (*Solanum nigrum*), naranjilla (*Solanum quitoense*), henbane (*Hyoscyamus niger*), and painted tongue (*Salpiglossis sinuata*). We investigated these publicly available large-scale datasets to determine if *SpKAR1* orthologs exhibit trichome-enriched expression. Phylogenetic analysis identified two distinct clades, and members of one clade are enriched in trichomes (Figure 4C;

Supplemental Figures S5 and S6). This suggests that SpKAR1 orthologs have a role in trichome acylsugar biosynthesis in other solanaceous species.

### Phylogenetic analysis of SpRBCS1

Rubisco is composed of eight large subunits (encoded by a single *RBCL* gene on the plastid genome) and eight small subunits (encoded by a multigene family *RBCS* on the nuclear genome). Among the five annotated *RBCS* genes in *S. pennellii* (three on chromosome 2 and one on each of chromosomes 3 and 7), only *Sopen07g006810* (*SpRBCS1*) exhibited noticeable trichome-enriched expression based on reverse transcription–quantitative PCR (RT–qPCR; Supplemental Figure S7A). In shaved stems, *SpRBCS1* had extremely low, if any, level of expression (if *SpRBCS1* is trichome specific, residual expression could be due to incomplete shaving of stem trichomes). On the other hand, other *RBCS* members showed high expression levels in shaved stems (approximately 1,000-fold higher than *SpRBCS1*). Additionally, protein sequence analysis revealed that other RBCS members, which share >90% sequence identity with each other, share a low level of identity (<55%) with SpRBCS1 (Supplemental Figure S7, A and B). Furthermore, *Sopen07g006810-Solyc07g017950* ortholog pair had dN/dS ratio of 2.91, whereas ortholog pairs of other RBCS members showed dN/dS ratios <0.32 (Supplemental Data Set 3), which is expected from highly conserved, mesophyll-expressed "regular" *RBCS* members (presumably involved in primary metabolism). These results indicated that SpRBCS1 is distinct from other RBCS members. This was supported by phylogenetic analysis, which in conjunction with publicly available large-scale dataset analysis, showed that orthologs of *SpRBCS1* are enriched in trichomes of solanaceous members (average 112-fold in 6 species where trichome-enriched expression data are available; Figure 5; Supplemental Figures S8 and S9). Interestingly, *SpRBCS1* was placed in a monophyletic clade with tobacco (*Nicotiana tabacum*) *RBCS-T*, which is expressed in trichome tip cells (Laterre et al., 2017). Orthologs of SpRBCS1 were found outside the Solanaceae (including *O. sativa*), but not in Arabidopsis (*Arabidopsis thaliana*).

### $\delta^{13}C$ analyses of acylsugars

Based on a previous metabolomics study (Balcke et al., 2017), it was hypothesized that Rubisco in trichomes mainly recycles $CO_2$ released by the high metabolic rate in these cells. In *S. pennellii* trichomes, where production of acylsugars is high, three steps in this pathway release $CO_2$: acetolactate synthase, isopropylmalate dehydrogenase, and branched-chain ketoacid dehydrogenase. If trichome Rubisco is responsible for recycling $CO_2$, then acylsugars (and other metabolites in trichomes) will contain some carbon that has undergone at least two rounds of fixation, first in the bulk of the plant to provide primary metabolites, and again in the trichome to recover $CO_2$ released during production of branched chain fatty acids. One consequence of this re-fixation would be that acylsugars contain even less $^{13}C$ than the rest of the plant due to isotopic fractionation at each fixation (the lighter $^{12}C$ isotope would be favored because of a lower activation energy). To test this hypothesis, we measured fractionation of carbon isotopes ($\delta^{13}C$; reported in parts per thousand ‰) in both secreted acylsugars and plant tissues presumably without acylsugars (shaved stems). Acylsugars and shaved stems showed $\delta^{13}C$ of −33.15‰ and −28.99‰, respectively (Figure 6A). The difference in $\delta^{13}C$ between these two sample types indicates that acylsugars contain carbons that have undergone additional rounds of fixation compared to nonacylsugar metabolites. Additionally, $\delta^{13}C$ values from *S. pennellii* samples (acylsugars and shaved stems) are comparable to $\delta^{13}C$ value from another member of the Solanaceae that was reported previously (−30.7‰ in tobacco) (Smith and Epstein, 1971).

Next, we hypothesized that the trichome-enriched *SpRBCS1* is responsible for this re-fixation of carbon into acylsugars. To test our hypothesis, we determined the difference in $\delta^{13}C$ between shaved stems and acylsugars ($\delta\delta^{13}C$) for two groups of plants—control and VIGS-*SpRBCS1*. Compared to a control group, VIGS led to a reduction in $\delta\delta^{13}C$ (Figure 6B), which confirmed the role of SpRBCS1 in supplying re-fixed carbon for acylsugar biosynthesis.

## Discussion

Acylsugars are secreted by glandular trichomes of several plant families, including Martyniaceae (Asai et al., 2010), Caryophyllaceae (Asai et al., 2012), Geraniaceae (Sakai et al., 2013), and Solanaceae, where they have been best studied. A detailed knowledge about genes involved in regulating acylsugar amount and acyl chain profile (Ben-Mahmoud et al., 2018) is required for successful crop breeding programs and metabolic engineering of acylsugar production. Here, we report two trichome-enriched genes involved in acylsugar biosynthesis.

### Multiple approaches to identify and validate candidate AMGs

Many enzymes involved in acylsugar biosynthesis have been identified in recent years (Ning et al., 2015; Fan et al., 2016, 2020; Moghe et al., 2017; Nadakuduti et al., 2017; Chang et al., 2020, 2022; Mandal et al., 2020; Feng et al., 2021; Lou et al., 2021; Schenck et al., 2022). To identify additional enzymes in this pathway, and hopefully regulatory proteins, we used a multipronged strategy to find genes whose expression and evolutionary patterns suggested they could be involved. To refine our previously reported list of candidate AMGs (Mandal et al., 2020), we looked for genes that were differentially expressed between high- and low-acylsugar-producing $F_2$ individuals derived from *S. lycopersicum* × *S. pennellii*, and also genes that appeared to be rapidly evolving as determined by the dN/dS ratio. The intersection of three sets of candidate AMGs yielded four candidates (Figure 2B). One candidate, *Sopen05g034770*, had lower expression in high-acylsugar-producing $F_2$ individuals than in low-acylsugar-producing $F_2$ individuals and also lower expression

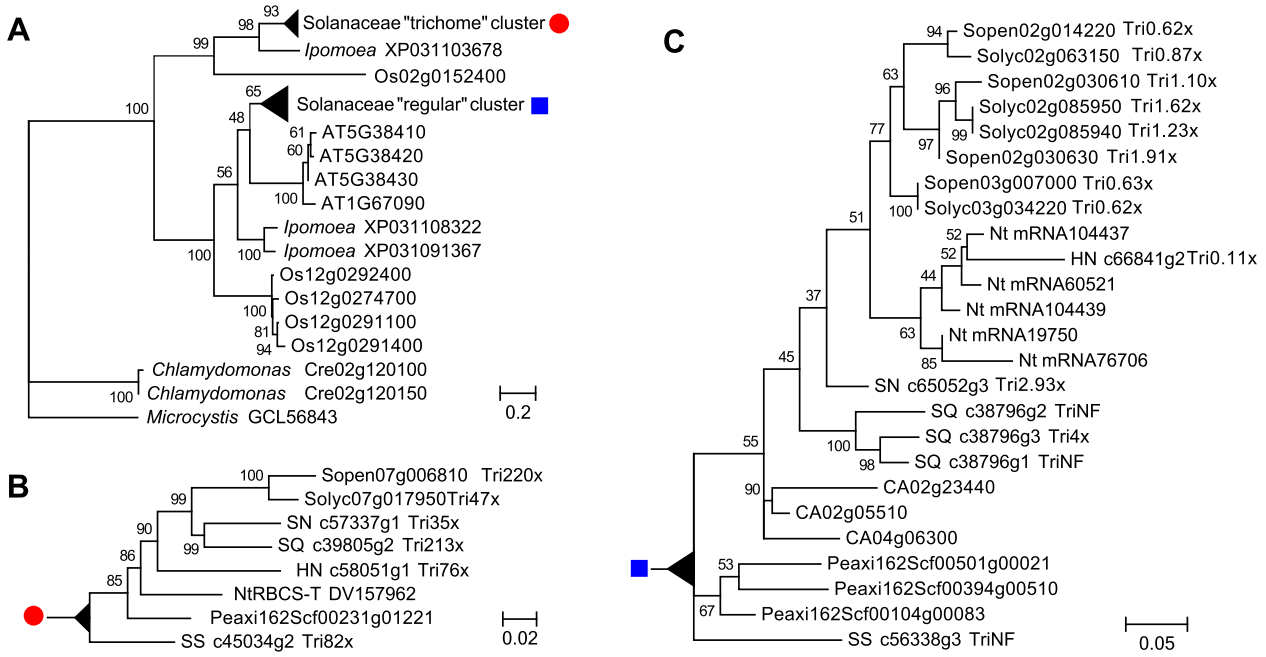


**Figure 5** Phylogenetic analysis of SpRBCS1 (Sopen07g006810; highlighted). Bootstrap values from 1,000 replicates are shown on the nodes. Trees are drawn to scale, with branch lengths measured in the number of substitutions per site. A, Maximum-likelihood tree of Rubisco small subunits. Solanaceous sequences were combined into two clusters (indicated by a red circle and a blue square) to save space. Neighbor-joining tree is given in Supplemental Figure S8. Os, *O. sativa*; AT, *A. thaliana*. GenBank accession numbers are indicated for *I. triloba* and *Microcystis aeruginosa*. B and C, Expanded Solanaceae "trichome" cluster (B) and expanded Solanaceae "regular" cluster (C). Tri220x indicates 220-fold higher expression in isolated trichomes compared to underlying tissues (NF, not found). RT-qPCR was used for *S. pennellii* (Sopen) sequences. Trichome-enriched expression data (based on RNA-seq) for sequences in five other species were obtained from Ning et al. (2015) (Solyc, *S. lycopersicum*) and Moghe et al. (2017) (SN, *S. nigrum*; SQ, *S. quitoense*; HN, *H. niger*; SS, *S. sinuata*). Peaxi, *P. axillaris*; CA, *C. annuum*; Nt, *N. tabacum*.

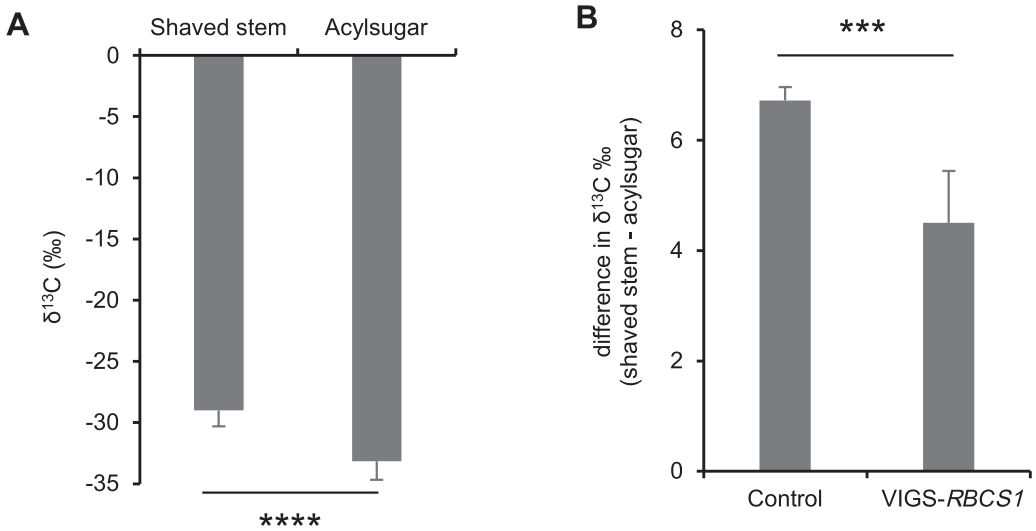


**Figure 6** $\delta^{13}C$ analyses. A, Difference in $\delta^{13}C$ values between shaved stems and secreted acylsugars. Error bars indicate SE ($n$ = 10 individual plants; ****$P < 0.0001$; Welch $t$ test). B, VIGS of *SpRBCS1* reduces the difference in $\delta^{13}C$ values between shaved stems and acylsugars. Error bars indicate SE ($n$ = 5 and 10 individual plants for control and VIGS-*RBCS1* groups, respectively; ***$P < 0.001$; Welch $t$ test).

in trichomes relative to underlying tissue (Table 1 and Figure 2C). Both of these patterns are opposite to what we would expect for genes involved in acylsugar biosynthesis. Therefore, this gene was not studied further, but the hypothetical protein encoded by this gene may act as a negative regulator of the pathway. Putative roles of the remaining three trichome-enriched candidates in acylsugar biosynthesis were tested using VIGS, which confirmed *SpKAR1* and *SpRBCS1* as AMGs (Figure 3). Phylogenetic analyses and publicly available large-scale dataset analyses revealed that orthologs of both *SpKAR1* (Figure 4C) and *SpRBCS1* (Figure 5) are trichome-enriched members in the

Solanaceae, suggesting roles for these orthologs in acylsugar biosynthesis. Additionally, *SpKAR1* and *SpRBCS1* exemplify duplication of highly-conserved primary metabolic genes followed by spatial regulation of gene expression as a driver of evolution of acylsugar metabolism.

## Function and phylogeny of SpKAR1

In plants, de novo fatty acid biosynthesis takes place predominantly in plastids, which use a multicomponent type II FAS system that catalyzes the extension of the growing acyl chain. The plastidic FAS has four well-characterized enzymatic components: a beta-ketoacyl-(acyl-carrier-protein) reductase (KAR), a beta-hydroxyacyl-(acyl-carrier-protein) dehydrase, an enoyl-(acyl-carrier-protein) reductase, and three isozymes of beta-ketoacyl-(acyl-carrier-protein) synthases (KAS I, II, and III). KAR catalyzes one of the steps of the core four-reaction cycle of the FAS-mediated chain elongation process. Close phylogenetic relationship with bacterial FabG (KAR) and its predicted chloroplast location indicate that SpKAR1 is a component of the plastid FAS, and it is distinct from other specialized metabolic cytosolic SDRs (Figure 4A). VIGS showed that SpKAR1 is required for acylsugar SCFA biosynthesis (Figure 3D), and we previously reported two trichome-enriched KAS genes that are involved in SCFA biosynthesis (Mandal et al., 2020). These results corroborate the role of FAS in acylsugar SCFA biosynthesis.

Elongation of acetyl-CoA to SCFAs by FAS complex presumably takes place in plastids; however, it is not clear if elongation of 2-methylpropanoyl-CoA (isobutyryl-CoA; produced in mitochondria from valine) to 8-methylnonanoyl-CoA (one of the major branched-chain acyl groups) occurs in mitochondria or chloroplast (Figure 4B). For both capsaicinoid (Mazourek et al., 2009) and acylsugar (Slocombe et al., 2008) biosynthesis, it has been proposed that 2-methylpropanoyl-CoA is transported from mitochondria to chloroplast, where it is elongated by the plastid FAS complex. Recently, a dual-localized KAR (AT1G24360) was reported in Arabidopsis (Guan et al., 2020), and AT1G24360 is phylogenetically closely related to SpKAR1 (Figure 4C; Supplemental Figure S5). This suggests the possibility that plastidic-SpKAR1 is involved in the biosynthesis of acylsugar SCFAs, whereas mitochondrial-SpKAR1 is involved in the elongation of 2-methylpropanoyl-CoA; this would allow 2-methylpropanoyl-CoA to be elongated without being transported to plastid.

It is likely that one-carbon elongation mechanism exists in acylsugars of Petunia (*Petunia axillaris*) and *S. sinuata*, based on information from Kroumova and Wagner (2003) and Moghe et al. (2017), respectively. Orthologs of SpKAR1 were found in these two species (Figure 4C). However, we observed a 15-aa deletion at the C-terminal end of the *P. axillaris* ortholog (Peaxi162Scf00128g00834) and two major changes at conserved sites for the *S. sinuata* ortholog (SS_c46569g1; Glu/Gln changed to His at position 132 and Leu/Ile changed to Phe at position 280) (Supplemental Figure S6). These changes may render these two orthologs nonfunctional. Alternatively, if these two orthologs are indeed functional, they may be involved in other pathways, such as elongation of epicuticular wax alkanes or other hydrocarbons.

## Capsaicinoid and acylsugar biosynthetic pathways

In *Capsicum*, 8-methyl-6-nonenoyl-CoA (C10) and 8-methylnonanoyl-CoA (C10) are attached to an aromatic compound derived from phenylalanine to generate capsaicin and dihydrocapsaicin, respectively, which are two of the most potent capsaicinoids (Mazourek et al., 2009). These C10 acyl molecules are presumably derived from 2-methylpropanoyl-CoA via FAS-mediated reactions, and this segment of the capsaicinoid biosynthetic pathway is similar to the acylsugar biosynthetic pathway (Figure 4B). Attachment of acyl groups to an aromatic compound (in case of capsaicinoid biosynthesis) or sucrose (in case of acylsugar biosynthesis) is catalyzed by BAHD superfamily transferases: PUN1 (for pungency) in capsaicinoid biosynthesis (Stewart et al., 2005) and ASATs in acylsugar biosynthesis (Schilmiller et al., 2015; Fan et al., 2016). Moghe et al. (2017) reported that PUN1 and ASATs share a common evolutionary ancestor, and our phylogenetic analysis revealed that SpKAR1 is closely related to two reductases involved in capsaicinoid biosynthesis (FAS-mediated acyl chain elongation steps) (Figure 4A). These results indicate close phylogenetic relationships between segments of capsaicinoid and acylsugar biosynthetic pathways. Additionally, acyl chain length is an important factor in determining both pungency of capsaicinoids and potential in defense activity of acylsugars (Ben-Mahmoud et al., 2018). Taken together, these findings will be useful in future metabolic engineering of capsaicinoid and acylsugar production.

## Function and phylogeny of SpRBCS1

VIGS results indicated that SpRBCS1 is involved in biosynthesis of both acylsugar and flavonoid (Figure 3), which are two most abundant classes of metabolites in trichomes of *S. pennellii* LA0716 (McDowell et al., 2011). Phylogenetic analysis clustered *SpRBCS1* with other solanaceous trichome-enriched *RBCS* members (Figure 5), including tobacco tip-cell-expressed *RBCS-T* (Laterre et al., 2017), and also with rice *OsRBCS1*, which is expressed in several tissues other than leaf blade (major photosynthetic organ) (Morita et al., 2014). Additionally, based on enzymatic properties, both RBCS-T and OsRBCS1 were found to be distinct from "regular" RBCS members in respective species. These findings suggest a broader role for this gene in specialized metabolism beyond acylsugar biosynthesis—trichome metabolism in the Solanaceae and specialized functions in other cell/tissue types outside the Solanaceae.

Trichome metabolites can accumulate at noticeably high levels; for example, acylsugars in *S. pennellii* and duvatriene-diol in tobacco can accumulate up to 20% and 15%, respectively, of leaf dry weight (Fobes et al., 1985; Severson et al., 1985). This indicates high metabolic activities and $CO_2$ release in trichome cells. For example, enzymatic steps

catalyzed by acetolactate synthase, isopropylmalate dehydrogenase, and branched-chain ketoacid dehydrogenase complex generate $CO_2$ during acylsugar production (each round of acyl chain elongation by FAS releases $CO_2$, but $CO_2$ is also consumed to generate malonyl-CoA from acetyl-CoA by acetyl-CoA carboxylase, leading to no net change in $CO_2$ during FAS-mediated elongation steps). However, due to thick cell walls and cuticle, it has been suggested that trichomes have limited gaseous exchange with the outside and fix little atmospheric $CO_2$ (Balcke et al., 2017). In order to support continued high metabolic activities, trichomes may require re-fixation of metabolic $CO_2$ with a Rubisco that is active in high $CO_2$ and low pH conditions. Biochemical assays indicate that tobacco RBCS-T has been adapted to such conditions (Laterre et al., 2017). These results encouraged us to investigate, using $\delta^{13}C$ analysis, if SpRBCS1 is involved in re-fixation of metabolic $CO_2$ into acylsugars.

Fractionation of carbon isotopes during photosynthesis occurs predominantly during the carboxylation reaction (carbon fixation) catalyzed by Rubisco, and it leads to a preferential enrichment of one stable isotope over another. Photosynthates contain less of the $^{13}C$ than the $^{12}C$ due to kinetic isotope effects—the lighter $^{12}C$ isotope is preferentially incorporated into products because it has a higher energy state (a lower activation energy) (Smith and Epstein, 1971). We used this information about carbon isotope fractionation to analyze metabolites, and tested the hypothesis based on Balcke et al. (2017) that in trichomes, Rubisco mostly re-fixes metabolic $CO_2$, which are incorporated into specialized metabolites (predominantly acylsugars in case of *S. pennellii*). Our $\delta^{13}C$ examination of acylsugars and shaved stems showed recycling of metabolic $CO_2$ into acylsugars (Figure 6A). Additionally, VIGS confirmed that the trichome-enriched *SpRBCS1* is responsible for this re-fixation (Figure 6B). Tomato trichomes receive carbon mostly from leaf sucrose (Balcke et al., 2017); however, re-fixation of metabolic $CO_2$ would allow trichomes to maintain a pH homeostasis and also to improve carbon utilization for sustained high-level production of specialized metabolites.

## Materials and methods

### Plant materials and growth conditions

Seeds of *S. pennellii* LA0716 and the $F_1$ hybrid (LA4135) of *S. lycopersicum* VF36 × *S. pennellii* LA0716 were obtained from the C.M. Rick Tomato Genetics Resource Center (University of California, Davis). The LA4135 was self-pollinated for generating the $F_2$ population. Seeds were treated with 1.2% (w/v) sodium hypochlorite for 20 min and rinsed with deionized water 3 times before placing on moist filter paper in petri dishes. After germination, seedlings were transferred to soil and grown in a growth chamber (16-h photoperiod; 24°C/20°C day/night temperature; 150 µMol $m^{-2}$ $s^{-1}$ photosynthetic photon flux density; 75% relative humidity).

### Acylsugar collection from $F_2$ individuals

Secreted acylsugars were collected from three young leaves of 10-week-old $F_2$ individuals as three replicates by dipping them in ethanol for 2–3 s. Ethanol was completely removed by evaporation until dryness in a fume hood. Acylsugar amount was determined as a proportion of leaf dry weight. Leaf dry weights were measured after drying in a 70°C oven for 1 week.

### RNA-seq

Ten high-acylsugar-producing $F_2$ individuals (acylsugar amounts > 14% of leaf dry weight) and 10 low-acylsugar-producing $F_2$ individuals (acylsugar amounts < 1% of leaf dry weight) were used in the HIGH-$F_2$ versus LOW-$F_2$ transcriptome comparison. After removing surface metabolites with ethanol for 2–3 s, young leaves were 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, Waltham, MA, USA), and the genomic DNA was removed using the TURBO DNA-Free Kit (Thermo Fisher Scientific). RNA-seq libraries of polyA$^+$-selected samples were prepared using TruSeq Stranded mRNA Library Preparation Kit LT (Illumina, San Diego, CA, USA). After quality control, libraries were sequenced on the HiSeq 4000 (Illumina) 150 × 150-bp paired-end sequencing platform according to the manufacturer's specifications at the Texas A&M Genomics and Bioinformatics Service Center, College Station.

### Differential gene expression analysis

Approximately 31–48 million (average 36 million) paired-end reads were generated from RNA-seq libraries. Sequencing reads were processed using Trimmomatic (Bolger et al., 2014a, 2014b). Approximately 68% of the reads in each library passed the trimming filter. The trimmed reads were then mapped to the *S. pennellii* LA0716 genome version 2.0 (Bolger et al., 2014a, 2014b) using TopHat2 (Kim et al., 2013). About 70%–85% of the trimmed reads were mapped to the *S. pennellii* genome. Aligned reads from TopHat2 were counted for each gene using HTseq (Anders et al., 2015). The count files were used to identify DEGs using edgeR version 3.32.1 (Robinson et al., 2010). Fragments per kilobase per million mapped reads value for each gene in each sample was called with rpkm command in edgeR program. Genes with more than one count per million in at least two samples were used for differential gene expression analysis. DEGs were identified when *P*, corrected for multiple testing, was < 0.05 (false discovery rate < 0.05), and fold change was > 2.

*Trimmomatic v0.32 parameters*: ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10, LEADING = 20, TRAILING = 20, SLIDINGWINDOW = 4:20, MINLEN = 100. *TopHat2 v2.1.0 parameters*: -library-type = fr-firststrand, -mate-inner-dist = 0, -mate-std-dev = 50, -read-realign-edit-dist = 1000, -read-edit-dist = 2, -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. *HTseq v 0.6.1 parameters*: -f bam, -r name, -s reverse, -m union, -a 20.

## Identification of putative orthologs and dN/dS estimation

To identify putative orthologs, we performed an all-versus-all reciprocal BLAST between annotated genes of *S. pennellii* version 2.0 (Bolger et al., 2014a, 2014b) and *S. lycopersicum* ITAG2.3 (Tomato Genome, 2012) with the following settings: minimum percentage identity = 70, minimum percentage query coverage = 50. Putative orthologs were aligned with ClustalW, and the alignment information was converted into codon alignments using PAL2NAL (Suyama et al., 2006). Genome-wide dN/dS ratios were calculated between putative ortholog pairs using the yn00 maximum likelihood method in the PAML package (Yang, 1997; Yang and Nielsen, 2000).

## Determination of trichome-enriched expression

RT–qPCR was used to measure trichome-enriched expression of selected genes in *S. pennellii* LA0716, as described in Mandal et al. (2020). RT–qPCR primers are given in Supplemental Table S1.

## VIGS

VIGS was performed using the TRV-based vectors (Dong et al., 2007) in *S. pennellii* LA0716. VIGS constructs were designed using the Solanaceae Genomics Network VIGS tool (http://vigs.solgenomics.net/) and were cloned into the pTRV2-LIC vector, in the antisense orientation, to target selected genes. *Agrobacterium tumefaciens* strain GV3101 harboring pTRV1, pTRV2 constructs, and empty pTRV2 were grown overnight with 50 mg $mL^{-1}$ kanamycin and 10 mg $mL^{-1}$ gentamicin at 28°C. Cultures were centrifuged at 8,000 *g* for 5 min at 4°C, and 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 different pTRV2 cultures were mixed with equal volumes of pTRV1 cultures to reach final $OD_{600}$ = 1 before infiltration at the first true leaf stage with a needleless syringe. Plants were grown in a chamber with conditions mentioned earlier for ~6 weeks. Silencing of target genes was assessed with RT–qPCR using primers that were designed outside the VIGS-targeted regions. VIGS primers are listed in Supplemental Table S1.

## Chromatography–mass spectrometry analysis

Secreted acylsugars from control and VIGS plants were quantified with LC–MS. Acylsugars on leaf surface were collected from similar-sized young leaves by submerging them in 10 mL of extraction solvent, followed by gentle mixing for 2 min. Extracted samples were analyzed using Q Exactive Focus coupled with Ultimate 3000 RS LC unit (Thermo Fisher Scientific) and Exactive Series 2.8 SP1/Xcalibur version 4.0 software. Acylsugars were separated by injecting 10 µL of sample into Acclaim 120 (2.1 × 150 mm; 3 µm) C18 column (Thermo Fisher Scientific) that was housed at 30°C. The Q Exactive Focus HESI source was operated in full MS in negative electrospray ionization mode. Parallel reaction monitoring mode was used for targeted MS/MS of acylsugars. Relative abundances of acylsugars were determined by dividing total peak areas of all detected acylsugars with peak area of the internal standard and leaf dry weight. Dry weights of the extracted leaves were measured after 1 week of drying in a 70°C oven.

*Extraction solvent* = acetonitrile:isopropanol:water (3:3:2, v/v/v) with 0.1% (v/v) formic acid; 100 µM propyl 4-hydroxybenzoate was used as the internal standard. *LC–MS parameters*: eluent A = 0.1% formic acid, eluent B = acetonitrile with 0.1% formic acid; flow rate = 300 µL $min^{-1}$ with the following gradient: (0–3 min, 40% B; 3–23 min, 40%–100% B; 23–28 min, hold 100% B; 28.1–31 min, hold 40% B); transfer capillary temperature and auxiliary gas heater temperature = 320°C and 350°C, respectively; sheath gas and auxiliary gas flow rates = 35 and 10 arbitrary units, respectively; spray voltage = 3.3 kV; S-Lens RF level = 50 V.

Acylsugar acyl chain profiles were analyzed with gas chromatography–mass spectrometry (GC–MS) after performing transesterification reaction, as described in Mandal et al. (2020).

## Phylogenetic analysis

Sequences were obtained from the Solanaceae Genomics Network, GenBank, and the NCBI website. For *S. nigrum*, *S. quitoense*, *H. niger*, and *S. sinuata*, de novo assembled transcriptomes (Moghe et al., 2017) were used to collect sequences, which were translated in six possible frames (https://web.expasy.org/translate/) to obtain protein sequences with longest open reading frames. MAFFT (Katoh and Standley, 2013) was used for multiple sequence alignment with BLOSUM62 matrix, gap extend penalty value of 0.123, and gap opening penalty value of 1.53. ModelFinder (Kalyaanamoorthy et al., 2017) was used to compare substitution models, and the best model of protein evolution was selected based on lowest Bayesian Information Criterion scores (LG + I + G4 for Figure 4A, JTT + G4 for Supplemental Figure S5, and LG + G4 for Figure 5). IQ-TREE 2 (Minh et al., 2020) was used to construct maximum likelihood-based phylogenetic trees. MEGA X (Kumar et al., 2018) was used to construct neighbor-joining phylogenetic trees applying Jones–Taylor–Thornton (JTT) model. Uniform rates among sites were used, and pairwise deletion was used for gaps/missing data. Bootstrap values were obtained from 1,000 replicates.

## Stable isotope analysis

$\delta^{13}C$ analysis was performed at the Texas A&M Stable Isotope Geosciences Facility using methods described in McDermott et al. (2019).

### Accession numbers

Sequence data from this article can be found in the GenBank data libraries under accession numbers ON920880 (SpRBCS1; Sopen07g006810), ON920881 (SpRBCS2; Sopen02g014220), ON920882 (SpRBCS3; Sopen02g030610), ON920883 (SpRBCS4; Sopen02g030630), ON920884 (SpRBCS5; Sopen03g007000), ON920885 (SpKAR1; Sopen05g009610), ON920886 (SpKAR2; Sopen06g028190), ON920887 (SpSTPL; Sopen05g032580), and ON920888 (hypothetical protein; Sopen05g034770). RNA-seq reads used in this study were submitted to the NCBI Sequence Read Archive under the accession number PRJNA818092 (BioProject ID).

## Supplemental data

The following materials are available in the online version of this article.

**Supplemental Figure S1.** Enrichment analysis of GO terms associated with 331 DEGs between high- and low-acylsugar-producing $F_2$ individuals.

**Supplemental Figure S2.** VIGS of candidate genes in *S. pennellii* LA0716.

**Supplemental Figure S3.** Multiple sequence alignment of SpKAR1 (Sopen05g009610) and its homologs, including specialized metabolic SDRs.

**Supplemental Figure S4.** Maximum-likelihood phylogenetic tree of SpKAR1 (Sopen05g009610) and related SDRs.

**Supplemental Figure S5.** Maximum-likelihood phylogenetic tree of SpKAR1 (Sopen05g009610) and its homologs in the Solanaceae.

**Supplemental Figure S6.** Multiple sequence alignment of SpKAR1 (Sopen05g009610) and its homologs in the Solanaceae.

**Supplemental Figure S7.** Analyses of RBCS members in *S. pennellii*.

**Supplemental Figure S8.** Neighbor-joining tree of SpRBCS1 (Sopen07g006810) and its homologs.

**Supplemental Figure S9.** Multiple sequence alignment of SpRBCS1 (Sopen07g006810) and its homologs.

**Supplemental Table S1.** List of primers used in this study.

**Supplemental Data Set 1.** DEGs between HIGH-$F_2$ and LOW-$F_2$ groups.

**Supplemental Data Set 2.** Reciprocal best hits between *S. pennellii* and *S. lycopersicum* annotated sequences.

**Supplemental Data Set 3.** 732 putative ortholog pairs between *S. pennellii* and *S. lycopersicum* with dN/dS $> 1$.

**Supplemental Data Set 4.** Acylsugars from *S. pennellii*.

## Acknowledgments

We sincerely thank Charlie Johnson and the staff of the Texas A&M Genomics and Bioinformatics Center for performing Illumina sequencing, Michael Dickens, and the staff of the Texas A&M High Performance Research Computing Platform for providing computational resources and assistance, Christopher Maupin of the Texas A&M Stable Isotope Geosciences Facility for assistance with $\delta^{13}C$ analysis, and the C.M. Rick Tomato Genetics Resource Center for supplying seeds.

## Funding

Early stages of this work were funded by a grant 2011-38821-30891 from the US Department of Agriculture.

*Conflict of interest statement.* None declared.

## References

**Alba JM, Montserrat M, 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

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

**Asai T, Hara N, Fujimoto Y** (2010) Fatty acid derivatives and dammarane triterpenes from the glandular trichome exudates of *Ibicella lutea* and *Proboscidea louisiana*. Phytochemistry **71**: 877–894

**Asai T, Nakamura Y, Hirayama Y, Ohyama K, Fujimoto Y** (2012) Cyclic glycolipids from glandular trichome exudates of *Cerastium glomeratum*. Phytochemistry **82**: 149–157

**Balcke GU, Bennewitz S, Bergau N, Athmer B, Henning A, Majovsky P, Jiménez-Gómez JM, Hoehenwarter W, 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

**Ben-Mahmoud S, Smeda JR, Chappell TM, Stafford-Banks C, Kaplinsky CH, Anderson T, Mutschler MA, Kennedy GG, Ullman DE** (2018) Acylsugar amount and fatty acid profile differentially suppress oviposition by western flower thrips, *Frankliniella occidentalis*, on tomato and interspecific hybrid flowers. PLoS One **13**: e0201583

**Bolger A, Scossa F, Bolger ME, Lanz C, Maumus F, Tohge T, Quesneville H, Alseekh S, Sorensen I, Lichtenstein G, et al.** (2014a) The genome of the stress-tolerant wild tomato species *Solanum pennellii*. Nat Genet **46**: 1034–1038

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

**Bonierbale MW, Plaisted RL, Pineda O, Tanksley SD** (1994) QTL analysis of trichome-mediated insect resistance in potato. Theor Appl Genet **87**: 973–987

**Chang A, Hu Z, Chen B, Vanderschuren H, Chen M, Qu Y, Yu W, Li Y, Sun H, Cao J, et al.** (2022) Characterization of trichome-specific BAHD acyltransferases involved in acylsugar biosynthesis in *Nicotiana tabacum*. J Exp Bot **73**: 3913–3928

**Chang AX, Chen B, Yang AG, Hu RS, Feng QF, Chen M, Yang XN, Luo CG, Li YY, Wang YY** (2020) The trichome-specific acetolactate synthase NtALS1 gene, is involved in acylsugar biosynthesis in tobacco (*Nicotiana tabacum* L.). Planta **252**: 13

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

**Fan P, Miller AM, Schilmiller AL, Liu X, Ofner I, Jones AD, Zamir D, Last RL** (2016) *In vitro* reconstruction and analysis of evolutionary variation of the tomato acylsucrose metabolic network. Proc Natl Acad Sci USA **113**: E239–E248

**Fan P, Wang P, Lou YR, Leong BJ, Moore BM, Schenck CA, Combs R, Cao P, Brandizzi F, Shiu SH, et al.** (2020) Evolution of a plant gene cluster in Solanaceae and emergence of metabolic diversity. eLife **9**: e56717

**Feng H, Acosta-Gamboa L, Kruse LH, Tracy JD, Chung SH, Nava Fereira AR, Shakir S, Xu H, Sunter G, Gore MA, et al.** (2021) Acylsugars protect *Nicotiana benthamiana* against insect herbivory and desiccation. Plant Mol Biol **109**: 505–522

**Fobes JF, Mudd JB, Marsden MP** (1985) Epicuticular lipid accumulation on the leaves of *Lycopersicon pennellii* (Corr.) D'Arcy and *Lycopersicon esculentum* Mill. Plant Physiol **77**: 567–570

**Guan X, Okazaki Y, Zhang R, Saito K, Nikolau BJ** (2020) Dual-localized enzymatic components constitute the Fatty acid synthase systems in mitochondria and plastids. Plant Physiol **183**: 517–529

**Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS** (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods **14**: 587–589

**Katoh K, Standley DM** (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol **30**: 772–780

**Kavanagh KL, Jornvall H, Persson B, Oppermann U** (2008) Medium- and short-chain dehydrogenase/reductase gene and protein families: the SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes. Cell Mol Life Sci **65**: 3895–3906

**Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL** (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol **14**: R36

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

**Krause ST, Liao P, Crocoll C, Boachon B, Forster C, Leidecker F, Wiese N, Zhao D, Wood JC, Buell CR, et al.** (2021) The biosynthesis of thymol, carvacrol, and thymohydroquinone in Lamiaceae proceeds via cytochrome P450s and a short-chain dehydrogenase. Proc Natl Acad Sci USA **118**: e2110092118

**Kroumova AB, Wagner GJ** (2003) Different elongation pathways in the biosynthesis of acyl groups of trichome exudate sugar esters from various solanaceous plants. Planta **216**: 1013–1021

**Kroumova AB, Zaitlin D, Wagner GJ** (2016) Natural variability in acyl moieties of sugar esters produced by certain tobacco and other Solanaceae species. Phytochem **130**: 218–227

**Kumar S, Stecher G, Li M, Knyaz C, Tamura K** (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol **35**: 1547–1549

**Laterre R, Pottier M, Remacle C, Boutry M** (2017) Photosynthetic trichomes contain a specific Rubisco with a modified pH-dependent activity. Plant Physiol **173**: 2110–2120

**Lawson DM, Lunde CF, Mutschler MA** (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 BM, D'Ambrosio DA, Chappell TM, Halitschke R, De Jong DM, Kessler A, Kennedy GG, Mutschler MA** (2016) Differential and synergistic functionality of acylsugars in suppressing oviposition by insect herbivores. PLoS One **11**: e0153345

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

**Lou YR, Anthony TM, Fiesel PD, Arking RE, Christensen EM, Jones AD, Last RL** (2021) It happened again: convergent evolution of acylglucose specialized metabolism in black nightshade and wild tomato. Sci Adv **7**: eabj8726

**Luu VT, Weinhold A, Ullah C, Dressel S, Schoettner M, Gase K, Gaquerel E, Xu S, Baldwin IT** (2017) *O*-acyl sugars protect a wild tobacco from both native fungal pathogens and a specialist herbivore. Plant Physiol **174**: 370–386

**Mandal S, Ji W, McKnight TD** (2020) Candidate gene networks for acylsugar metabolism and plant defense in wild tomato *Solanum pennellii*. Plant Cell **32**: 81–99

**Mazourek M, Pujar A, Borovsky Y, Paran I, Mueller L, Jahn MM** (2009) A dynamic interface for capsaicinoid systems biology. Plant Physiol **150**: 1806–1821

**McDermott EG, Mullens BA, Mayo CE, Roark EB, Maupin CR, Gerry AC, Hamer GL** (2019) Laboratory evaluation of stable isotope labeling of Culicoides (Diptera: Ceratopogonidae) for adult dispersal studies. Parasit Vect **12**: 411

**McDowell ET, Kapteyn J, Schmidt A, Li C, Kang JH, Descour A, Shi F, Larson M, Schilmiller A, An L, et al.** (2011) Comparative functional genomic analysis of *Solanum* glandular trichome types. Plant Physiol **155**: 524–539

**Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R** (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol **37**: 1530–1534

**Moghe GD, Last RL** (2015) Something old, something new: enzymes and the evolution of novelty in plant specialized metabolism. Plant Physiol **169**: 1512–1523

**Moghe GD, Leong BJ, Hurney SM, Daniel Jones A, Last RL** (2017) Evolutionary routes to biochemical innovation revealed by integrative analysis of a plant-defense related specialized metabolic pathway. eLife **6**: e28468

**Moore BM, Wang P, Fan P, Leong B, Schenck CA, Lloyd JP, Lehti-Shiu MD, Last RL, Pichersky E, Shiu SH** (2019) Robust predictions of specialized metabolism genes through machine learning. Proc Natl Acad Sci USA **116**: 2344–2353

**Morita K, Hatanaka T, Misoo S, Fukayama H** (2014) Unusual small subunit that is not expressed in photosynthetic cells alters the catalytic properties of Rubisco in rice. Plant Physiol **164**: 69–79

**Nadakuduti SS, Uebler JB, Liu X, Jones AD, Barry CS** (2017) Characterization of trichome-expressed BAHD acyltransferases in *Petunia axillaris* reveals distinct acylsugar assembly mechanisms within the Solanaceae. Plant Physiol **175**: 36–50

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

**Okamoto S, Yu F, Harada H, Okajima T, Hattan J, Misawa N, Utsumi R** (2011) A short-chain dehydrogenase involved in terpene metabolism from *Zingiber zerumbet*. FEBS J **278**: 2892–2900

**Polichuk DR, Zhang Y, Reed DW, Schmidt JF, Covello PS** (2010) A glandular trichome-specific monoterpene alcohol dehydrogenase from *Artemisia annua*. Phytochemistry **71**: 1264–1269

**Ringer KL, Davis EM, Croteau R** (2005) Monoterpene metabolism. Cloning, expression, and characterization of (-)-isopiperitenol/(-)--carveol dehydrogenase of peppermint and spearmint. Plant Physiol **137**: 863–872

**Robinson MD, McCarthy DJ, Smyth GK** (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26**: 139–140

**Sakai T, Tanemura Y, Itoh S, Fujimoto Y** (2013) Dodecyl alpha-L-rhamnopyranosyl-(1– > 2)-beta-D-fucopyranoside derivatives from the glandular trichome exudate of *Erodium pelargoniflorum*. Chem Biodivers **10**: 1099–1108

**Schenck CA, Anthony TM, Jacobs M, Jones AD, Last RL** (2022) Natural variation meets synthetic biology: promiscuous trichome-Expressed acyltransferases from *Nicotiana*. Plant Physiol **190**: 146–164

**Schilmiller AL, Moghe GD, Fan P, Ghosh B, Ning J, Jones AD, Last RL** (2015) Functionally divergent alleles and duplicated loci encoding an acyltransferase contribute to acylsugar metabolite diversity in *Solanum* trichomes. Plant Cell **27**: 1002–1017

**Severson RF, Johnson AW, Jackson DM** (1985) Cuticular constituents of tobacco: factors affecting their production and their role in insect and disease resistance and smoke quality. Recent Adv Tobacco Sci **11**: 105–173

**Shimura K, Okada A, Okada K, Jikumaru Y, Ko KW, Toyomasu T, Sassa T, Hasegawa M, Kodama O, Shibuya N, et al.** (2007)

Identification of a biosynthetic gene cluster in rice for momilactones. J Biol Chem **282**: 34013–34018

**Slocombe SP, Schauvinhold I, McQuinn RP, Besser K, Welsby NA, Harper A, Aziz N, Li Y, Larson TR, Giovannoni J, et al.** (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

**Smith BN, Epstein S** (1971) Two categories of c/c ratios for higher plants. Plant Physiol **47**: 380–384

**Sonawane PD, Heinig U, Panda S, Gilboa NS, Yona M, Kumar SP, Alkan N, Unger T, Bocobza S, Pliner M, et al.** (2018) Short-chain dehydrogenase/reductase governs steroidal specialized metabolites structural diversity and toxicity in the genus *Solanum*. Proc Natl Acad Sci USA **115**: E5419–E5428

**Stewart C, Jr., Kang BC, Liu K, Mazourek M, Moore SL, Yoo EY, Kim BD, Paran I, Jahn MM** (2005) The Pun1 gene for pungency in pepper encodes a putative acyltransferase. Plant J **42**: 675–688

**Suyama M, Torrents D, Bork P** (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res **34**: W609–612

**Tomato Genome C** (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature **485**: 635–641

**Walters DS, Steffens JC** (1990) Branched chain amino acid metabolism in the biosynthesis of *Lycopersicon pennellii* glucose esters. Plant Physiol **93**: 1544–1551

**Weinhold A, Baldwin IT** (2011) Trichome-derived *O*-acyl sugars are a first meal for caterpillars that tags them for predation. Proc Natl Acad Sci USA **108**: 7855–7859

**Yang Z** (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci **13**: 555–556

**Yang Z, Nielsen R** (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol **17**: 32–43