Evolution of a plant gene cluster in Solanaceae and emergence of metabolic diversity

Pengxiang Fan[1], Peipei Wang[2], Yann-Ru Lou[1], Bryan J. Leong[2], Bethany M. Moore[2], Craig A. Schenck[1], Rachel Combs[3†], Pengfei Cao[2,4], Federica Brandizzi[2,4], Shin-Han Shiu[2,5], Robert L. Last[1,2*]

[1]*Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824*

[2]*Department of Plant Biology, Michigan State University, East Lansing, MI 48824*

[3]*Division of Biological Sciences, University of Missouri, Columbia, MO 65211*

[4]*MSU-DOE Plant Research Laboratory, Michigan State University, East Lansing, United States, East Lansing, MI 48824*

[5]*Department of Computational Mathematics, Science, and Engineering, Michigan State University, East Lansing, MI 48824*

[†]*Current address: Center for Applied Plant Sciences, The Ohio State University, Columbus, OH 43210*

**Corresponding Author: lastr@msu.edu*

Short title: Plant Metabolic Evolution

Abstract (<150 words)

Plants produce phylogenetically and spatially restricted, as well as structurally diverse specialized metabolites via multistep metabolic pathways. Hallmarks of specialized metabolic evolution include enzymatic promiscuity and recruitment of primary metabolic enzymes and examples of genomic clustering of pathway genes. Solanaceae glandular trichomes produce defensive acylsugars, with sidechains that vary in length across the family. We describe a tomato gene cluster on chromosome 7 involved in medium chain acylsugar accumulation due to trichome specific acyl-CoA synthetase and enoyl-CoA hydratase genes. This cluster co-localizes with a tomato steroidal alkaloid gene cluster and is syntenic to a chromosome 12 region containing another acylsugar pathway gene. We reconstructed the evolutionary events leading to this gene cluster and found that its phylogenetic distribution correlates with medium chain acylsugar

accumulation across the Solanaceae. This work reveals insights into the dynamics behind gene cluster evolution and cell-type specific metabolite diversity.

Introduction

Despite the enormous structural diversity of plant specialized metabolites, they are derived from a relatively small number of primary metabolites, such as sugars, amino acids, nucleotides, and fatty acids (Maeda, 2019). These lineage-, tissue- or cell- type specific specialized metabolites mediate environmental interactions, such as herbivore and pathogen deterrence or pollinator and symbiont attraction (Mithöfer and Boland, 2012; Pichersky and Lewinsohn, 2011). Specialized metabolism evolution is primarily driven by gene duplication (Moghe and Last, 2015; Panchy et al., 2016), and relaxed selection of the resulting gene pairs allows modification of cell- and tissue-specific gene expression and changes in enzymatic activity. This results in expanded substrate recognition and/or diversified product formation (Khersonsky and Tawfik, 2010; Leong and Last, 2017). The neofunctionalized enzymes can prime the origin and diversification of specialized metabolic pathways (Schenck and Last, 2019; Weng et al., 2012; Weng, 2014).

There are many examples of mechanisms that lead to novel enzymatic activities in specialized cell- or tissue-types, however, the principles that govern assembly of multi-enzyme specialized metabolic pathways are less well established. One appealing hypothesis involves the stepwise recruitment of pathway enzymes (Noda-Garcia et al., 2018). In rare cases, non-homologous specialized metabolic enzyme genes occur in proximity to each other in a genomic region, forming a biosynthetic gene cluster (Nützmann et al., 2016; Nützmann and Osbourn, 2014; Rokas et al., 2018). In recent years, an increasing number of specialized metabolic gene clusters (SMGCs) were experimentally identified or bioinformatically predicted in plants (Boutanaev et al., 2015; Castillo et al., 2013; Schläpfer et al., 2017). However, although most experimentally

54 characterized plant SMGCs are co-expressed, the majority of the bioinformatically predicted ones
55 do not show coexpression under global network analysis (Wisecaver et al., 2017).

56 While examples of SMGCs are still relatively rare in plants, experimentally validated
57 cases were reported for a surprisingly diverse group of pathways. These include terpenes (Chae et
58 al., 2014; Prisic et al., 2004; Qi et al., 2004; Wilderman et al., 2004), cyclic hydroxamic acids
59 (Frey et al., 1997), biosynthetically unrelated alkaloids (Itkin et al., 2013; Winzer et al., 2012),
60 polyketides (Schneider et al., 2016), cyanogenic glucosides (Takos et al., 2011), and modified
61 fatty acids (Jeon et al., 2020). However, whereas each cluster encodes multiple non-homologous
62 enzymes of a biosynthetic pathway, evolution of their assembly is not well understood.

63 Acylsugars are a group of insecticidal (Leckie et al., 2016) and anti-inflammatory
64 (Herrera-Salgado et al., 2005) chemicals mainly observed in glandular trichomes of Solanaceae
65 species (Fan et al., 2019; Schuurink and Tissier, 2019). These specialized metabolites are sugar
66 aliphatic esters with three levels of structural diversity across the Solanaceae family: acyl chain
67 length, acylation position, and sugar core (Fan et al., 2019). The primary metabolites sucrose and
68 aliphatic acyl-CoAs are the biosynthetic precursors of acylsucroses in plants as evolutionarily
69 divergent as the cultivated tomato *Solanum lycopersicum* (Fan et al., 2016) (Figure 1), *Petunia*
70 *axillaris* (Nadakuduti et al., 2017) and *Salpiglossis sinuata* (Moghe et al., 2017). The core tomato
71 acylsucrose biosynthetic pathway involves four BAHD [**B**EAT, **A**HCT, **H**CBT, **D**AT (D'Auria,
72 2006)] family acylsucrose acyltransferases (*Sl-ASAT1* through *Sl-ASAT4*), which are specifically
73 expressed in the type I/IV trichome tip cells (Fan et al., 2016; Schilmiller et al., 2015, 2012).
74 These enzymes catalyze consecutive reactions utilizing sucrose and acyl-CoA substrates to
75 produce the full set of cultivated tomato acylsucroses *in vitro* (Fan et al., 2016).

76 Co-option of primary metabolic enzymes contributed to the evolution of acylsugar
77 biosynthesis and led to interspecific structural diversification across *Solanum* tomato clade. One
78 example is an invertase-like enzyme originating from carbohydrate metabolism that generates

79 acylglucoses in the wild tomato *S. pennellii* through cleavage of the acylsucrose glycosidic bond
80 (Leong et al., 2019). In another case, allelic variation of a truncated isopropylmalate synthase-like
81 enzyme (IPMS3) – from branched chain amino acid metabolism – leads to acylsugar iC4/iC5 (2-
82 methylpropanoic/3-methylbutanoic acid) acyl chain diversity in *S. pennellii* and *S. lycopersicum*
83 (Ning et al., 2015). Acylsugar structural diversity is even more striking across the family.
84 Previous studies revealed variation in acyl chain length (Ghosh et al., 2014; Liu et al., 2017;
85 Moghe et al., 2017): *Nicotiana*, *Petunia* and *Salpiglossis* species were reported to accumulate
86 acylsugars containing only short acyl chains (carbon number, C≤8). In contrast, some species in
87 *Solanum* and other closely related genera produce acylsugars with medium acyl chains (C≥10).
88 These results are consistent with the hypothesis that the capability to produce medium chain
89 acylsugars varies across the Solanaceae family.

90 In this study, we identify a metabolic gene cluster on tomato chromosome 7 containing
91 two non-homologous genes – acylsugar acyl-CoA synthetase (*AACS*) and acylsugar enoyl-CoA
92 hydratase (*AECH*) – affecting medium chain acylsugar biosynthesis. Genetic and biochemical
93 results show that the trichome enriched *AACS* and *AECH* are involved in generating medium
94 chain acyl-CoAs, which are donor substrates for acylsugar biosynthesis. Genomic analysis
95 revealed a syntenic region on chromosome 12, where the acylsucrose biosynthetic *Sl-ASAT1* is
96 located (Fan et al., 2016). Phylogenetic analysis of the syntenic regions in Solanaceae and beyond
97 led to evolutionary reconstruction of the origin of the acylsugar gene cluster. We infer that
98 sequential gene insertion facilitated emergence of this gene cluster in tomato. These results
99 provide insights into specialized metabolic evolution through emergence of cell-type specific
100 gene expression, the formation of metabolic gene clusters and illuminates additional examples of
101 primary metabolic enzymes being co-opted into specialized metabolism.

102 **Results**

103 **Identification of a metabolic gene cluster that affects tomato trichome medium chain**
104 **acylsugar biosynthesis**

105 *S. pennellii* natural accessions (Mandal et al., 2020), as well as the *S. lycopersicum* M82 × *S.*
106 *pennellii* LA0716 chromosomal substitution introgression lines (ILs) (Eshed and Zamir, 1995),
107 offer convenient resources to investigate interspecific genetic variation that affects acylsugar
108 metabolic diversity (Mandal et al., 2020; Schilmiller et al., 2010). In a rescreen of ILs for *S.*
109 *pennellii* genetic regions that alter trichome acylsugar profiles (Schilmiller et al., 2010), IL7-4
110 was found to accumulate increased C10 medium chain containing acylsugars compared with M82
111 (Figure 2, A and B). The genetic locus that contributes to the acylsugar phenotype was narrowed
112 down to a 685 kb region through screening selected backcross inbred lines (BILs) (Ofner et al.,
113 2016) that have recombination breakpoints on chromosome 7 (Figure 2C). Because tomato
114 acylsucrose biosynthesis occurs in trichomes, candidate genes in this region were filtered based
115 on their trichome-specific expression patterns. This analysis identified a locus containing multiple
116 tandemly duplicated genes of three families – an acyl-CoA synthetase (ACS), enoyl-CoA
117 hydratase (ECH), and BAHD acyltransferase. Our analysis (Moore et al., 2020) revealed co-
118 expression of four *Sl-ASATs* (Fan et al., 2019) and three genes at the locus – *Solyc07g043630*,
119 *Solyc07g043660*, and *Solyc07g043680* (Supplementary file1 and Figure 2 — figure supplement
120 1). Expression of these three genes was trichome enriched (Figure 2D), and thus they were
121 selected for further analysis.

122 The three candidate genes were tested for involvement in tomato acylsugar biosynthesis
123 by making loss of function mutations using the CRISPR-Cas9 gene editing system. Two guide
124 RNAs (gRNAs) were designed to target one or two exons of each gene to assist site-specific DNA
125 cleavage by hCas (Brooks et al., 2014) (Figure 3 — figure supplement 1, A-C). In the self-crossed
126 T1 progeny of stably transformed M82 plants, at least two homozygous mutants were obtained in
127 *Solyc07g043630*, *Solyc07g043660*, and *Solyc07g043680* (Figure 3 — figure supplement 1, A-C),

128 and these were analyzed for leaf trichome acylsugar changes. Altered acylsugar profiles were
129 observed in the ACS-annotated *Solyc07g043630* or ECH-annotated *Solyc07g043680* mutants
130 (Figure 3, A and B), but not in the ACS-annotated *Solyc07g043660* mutant (Figure 3 — figure
131 supplement 1D). Despite carrying mutations in distinctly annotated genes (ACS or ECH), the two
132 mutants exhibited the same phenotype – no detectable medium acyl chain (C10 or C12)
133 containing acylsugars (Figure 3, A and B). We renamed *Solyc07g043630* as *acylsugar acyl-CoA*
134 *synthetase 1* (*Sl-AACS1*) and *Solyc07g043680* as *acylsugar enoyl-CoA hydratase 1* (*Sl-AECH1*)
135 based on this analysis.

136 Further genomic analysis revealed that *Sl-AACS1* and *Sl-AECH1* belong to a syntenic region
137 shared with a locus on chromosome 12, where *Sl-ASAT1* is located (Figure 3C and Figure 3 —
138 figure supplement 2). *Sl-ASAT1* is specifically expressed in trichome tip cells and encodes the
139 enzyme catalyzing the first step of tomato acylsucrose biosynthesis (Fan et al., 2016). This led us
140 to test the cell-type expression pattern of *Sl-AACS1* and *Sl-AECH1*. Like *Sl-ASAT1*, the promoters
141 of both genes drove GFP expression in the trichome tip cells of stably transformed M82 plants
142 (Figure 3C). This supports our hypothesis that *Sl-AACS1* and *Sl-AECH1* are involved in tomato
143 trichome acylsugar biosynthesis. Taken together, we identified a metabolic gene cluster involved
144 in medium chain acylsugar biosynthesis, which is composed of two cell-type specific genes.

145 ***In vitro* analysis of Sl-AACS1 and Sl-AECH1 implicates their roles in medium chain acyl-**
146 **CoA metabolism**

147 ACS and ECH are established to function in multiple cell compartments for the metabolism of
148 acyl-CoA (Buchanan, B. B. Gruissem, W. Jones, 2015), the acyl donor substrates for ASAT
149 enzymes. We sought to understand the organelle targeting of Sl-AACS1 and Sl-AECH1, to
150 advance our knowledge of acylsugar machinery at the subcellular level. We constructed
151 expression cassettes of Sl-AACS1, Sl-AECH1 and Solyc07g043660 with C-terminal cyan
152 fluorescent protein (CFP), hypothesizing that the targeting peptides reside at the N-terminus of

153 precursor proteins. When co-expressed in tobacco leaf epidermal cells, three CFP-tagged
154 recombinant proteins co-localized with the mitochondrial marker MT-RFP (Nelson et al., 2007)
155 (Figure 4A and Figure 4 — figure supplement 1A). To rule out the possibility of peroxisomal
156 localization, we fused Sl-AACS1, Sl-AECH1, or Solyc07g043660 with N-terminus fused yellow
157 fluorescent protein (YFP), considering that potential peroxisomal targeting peptides are usually
158 located on the C-terminus (Brocard and Hartig, 2006). The expressed YFP-recombinant proteins
159 were not co-localized with the peroxisomal marker RFP-PTS (Nelson et al., 2007) (Figure 4 —
160 figure supplement 1B). Instead, they appeared distributed in the cytosol (Figure 4 — figure
161 supplement 1B), presumably because the N-terminal YFP blocked the mitochondria targeting
162 signal. Taken together, protein expression and co-localization analyses suggest that *Sl-AACS1*, *Sl-*
163 *AECH1,* and *Solyc07g043660* encode enzymes targeted to mitochondria.

164 Sl-AACS1 belongs to a group of enzymes that activate diverse carboxylic acid substrates to
165 produce acyl-CoAs. We hypothesized that Sl-AACS1 uses medium chain fatty acids as substrates,
166 because ablation of *Sl-AACS1* eliminated acylsugars with medium acyl chains. To characterize the
167 *in vitro* activity of Sl-AACS1, we purified recombinant His-tagged proteins from *Escherichia coli*.
168 Enzyme assays were performed by supplying fatty acid substrates with even carbon numbers from
169 C2 through C18 (Figure 4B). The results showed that Sl-AACS1 utilized fatty acid substrates
170 with lengths ranging from C6 to C12, including those with a terminal branched carbon (iC10:0) or
171 an unsaturated bond (*trans*-2-decenoic acid, C10:1) (Figure 4, B and C). However, no activity
172 was observed with the 3-hydroxylated C12 and C14 fatty acids as substrates (Figure 4B). These
173 results support our hypothesis that Sl-AACS1 produces medium chain acyl-CoAs, which are *in*
174 *vivo* substrates for acylsugar biosynthesis.

175 To test whether *Sl-AACS1* and *Sl-AECH1* can produce medium chain acyl-CoAs *in planta,*
176 we transiently expressed these genes in *Nicotiana benthamiana* leaves using *Agrobacterium-*
177 mediated infiltration (Sainsbury et al., 2009). It is challenging to directly measure plant acyl-

178 CoAs, due to their low concentration and separate organellar pools. We used an alternative
179 approach and characterized membrane lipids, which are produced from acyl-CoA intermediates.
180 We took advantage of the observation that *N. benthamiana* membrane lipids do not accumulate
181 detectable acyl chains of 12 carbons or shorter. *N. benthamiana* leaves were infiltrated with
182 constructs containing *Sl-AACS1* or *Sl-AECH1* individually, or together (Figure 4D). In contrast to
183 the empty vector control, infiltration of *Sl-AECH1* led to detectable levels of C12 acyl chains in
184 the leaf membrane lipid phosphatidylcholine (PC) (Figure 4D). We also observed increased C14
185 acyl chains in PC, phosphatidylglycerol (PG), sulfoquinovosyl diacylglycerol (SQDG), and
186 digalactosyldiacylglycerol (DGDG) in *Sl-AECH1* infiltrated plants (Figure 4D and Figure 4 —
187 figure supplement 1C). These results suggest that *Sl-AECH1* participates in generation of medium
188 chain acyl-CoAs *in planta*, which are channeled into lipid biosynthesis. No medium chain
189 acylsugars were detected, presumably due to the lack of core acylsugar biosynthetic machinery in
190 *N. benthamiana* mesophyll cells.

191 We asked whether the closest known homologs of *Sl-AECH1* from *Solanum* species can
192 generate medium chain lipids when transiently expressed in *N. benthamiana*. Two SQDGs with
193 C12 chains were monitored by LC/MS as peaks diagnostic of lipids containing medium chain
194 fatty acids (Figure 4 — figure supplement 2, A and B). The results showed that only the putative
195 *Sl-AECH1* orthologs *Sopen07g023250* (*Sp-AECH1*) and *Sq_c37194* (*Sq-AECH1*) – from *S.*
196 *pennellii* and *S. quitoense* respectively – generated medium chain lipids in the infiltrated leaves
197 (Figure 4 — figure supplement 2C). This confirms that not all ECHs can produce medium chain
198 lipids and suggests that the function of *Sl-AECH1* evolved recently, presumably as a result of
199 neofunctionalization after gene duplication (Figure 4 — figure supplement 2C).

200 **_AACS1_ and _AECH1_ are evolutionarily conserved in the _Solanum_**

201 Medium chain acylsugars were documented in *Solanum* species besides cultivated tomato,
202 including *S. pennellii* (Leong et al., 2019), *S. nigrum* (Moghe et al., 2017), as well as the more

203 distantly related *S. quitoense* (Hurney, 2018) and *S. lanceolatum* (Herrera-Salgado et al., 2005).
204 We hypothesized that evolution of *AACS1* and *AECH1* contributed to medium chain acylsugar
205 biosynthesis in *Solanum.* As a test, we analyzed the genomes of *Solanum* species other than
206 cultivated tomato. Indeed, the acylsugar related synteny containing ACS and ECH was found in
207 both *S. pennellii* and *S. melongena* (eggplant), suggesting that the cluster assembly evolved before
208 divergence of the tomato and eggplant lineage (Figure 5A).

209 We applied gene expression and genetic approaches to test the *in vivo* functions of ACS and
210 ECH in selected *Solanum* species. To explore the expression pattern of *S. pennellii* ACS and ECH
211 cluster genes, we performed RNA-seq analysis on trichomes and shaved stems to identify
212 acylsugar biosynthetic candidates (Supplementary file2). The expression pattern of *S. pennellii*
213 cluster genes is strikingly similar to *S. lycopersicum*: one ECH and two ACS genes are highly
214 enriched in trichomes, including the orthologs of *Sl-AACS1* and *Sl-AECH1*. *Sp-AACS1* function
215 (*Sopen07g023200*) was first tested by asking whether it can reverse the cultivated tomato *sl-aacs1*
216 mutant acylsugar phenotype. Indeed, *Sp-AACS1* restored C12 containing acylsugars in the stably
217 transformed *sl-aacs1* plants (Figure 5 — figure supplement 1). To directly test *Sp-AACS1* and *Sp-*
218 *AECH1* function, we used CRISPR-Cas9 to make single mutants in *S. pennellii* LA0716. No
219 medium chain acylsugars were detected in T0 generation mutants with edits for each gene (Figure
220 5B and Figure 5 — figure supplement 2, A and C). Similar to the ACS-annotated
221 *Solyc07g043660* cultivated tomato mutant (Figure 3 — figure supplement 1D), deletion of *S.*
222 *pennellii* ortholog *Sopen07g023220* has no observed effects on *S. pennellii* trichome acylsugars
223 (Figure 5 — figure supplement 2D).

224 The medium chain acylsugar producer *S. quitoense* (Hurney, 2018) was used for *AACS1* and
225 *AECH1* functional analysis because of its phylogenetic distance from the tomato clade, it is in the
226 *Solanum* Leptostemonum clade (including eggplant), and the fact that it produces medium chain
227 acylsugars. We found trichome-enriched putative orthologs of *AACS1* and *AECH1* in the

228 transcriptome dataset of *S. quitoense* (Moghe et al., 2017), and tested their *in vivo* function
229 through virus-induced gene silencing (VIGS) (Figure 5 — figure supplement 2E). Silencing either
230 gene led to decreased total acylsugars (Figure 5C and Figure 5 — figure supplement 2F), which
231 correlated with the degree of expression reduction in each sample (Figure 5D). These results are
232 consistent with the hypothesis that *Sq-AACS1* and *Sq-AECH1* are involved in medium chain
233 acylsugar biosynthesis, because all acylsugars in *S. quitoense* carry two medium chains (Hurney,
234 2018). The importance of *AACS1* and *AECH1* in medium chain acylsugar biosynthesis in distinct
235 *Solanum* clades inspired us to explore the evolutionary origins of the gene cluster.

236 **Evolution of the gene cluster correlates with the distribution of medium chain acylsugars**
237 **across Solanaceae**

238 We sought to understand how the acylsugar gene cluster evolved and whether it correlates with
239 the distribution of medium chain acylsugars across the Solanaceae family. Taking advantage of
240 the available genome sequences of 13 species from Solanaceae and sister families, we analyzed
241 the regions that are syntenic with the tomato acylsugar gene cluster (Figure 6 — figure
242 supplement 1). This synteny was found in all these plants, including the most distantly related
243 species analyzed, *Coffea canephora* (coffee, Rubiaceae) (Figure 6 — figure supplement 1).
244 BAHD acyltransferases were the only genes observed in the syntenic regions both inside and
245 outside the Solanaceae, in contrast to ECH and ACS, which are restricted to the family (Figure
246 6A and Figure 6 — figure supplement 1). Within the syntenic regions of the species analyzed,
247 ECH homologs, including pseudogenes, are present in all Solanaceae except for *Capsicum*
248 species, while ACS is more phylogenetically restricted, being found only in *Nicotiana* and
249 *Solanum* (Figure 6A and Figure 6 — figure supplement 1).

250 We then performed phylogenetic analysis to reconstruct the evolutionary history of the
251 ACS, ECH, and BAHD acyltransferase genes in the syntenic region (Figure 6). This analysis

252 revealed a model for the temporal order of emergence for the three types of genes, leading to their
253 presence in the syntenic regions in extant Solanaceae plants (Figure 6B). We propose that the
254 BAHD acyltransferase was the first of three genes that emerged before the divergence between
255 Solanaceae and Rubiaceae, and was likely lost in Convolvulaceae. This hypothesis is based on the
256 discovery of a BAHD acyltransferase pseudogene in the syntenic region of *C. canephora* (Figure
257 6A and Figure 6 — figure supplement 1), which is one of the closest *Coffea* sequences sister to
258 the ASAT clade (Figure 6 — figure supplement 2 and Figure 6 — figure supplement 5). In our
259 model, ECH was likely inserted into the syntenic region before the Solanaceae-specific whole
260 genome duplication (WGD) event (Figure 6B and Figure 6 — figure supplement 3).

261 We propose that ACS was inserted into the synteny through segmental duplication (Bailey
262 et al., 2002) (Figure 6 — figure supplement 4). However, whether ACS insertion happened before
263 or after the Solanaceae-specific WGD event cannot be resolved by the phylogenetic analysis
264 (Figure 6 — figure supplement 4). If the insertion happened before WGD, one ACS gene loss on
265 chromosome 12 in *Solanum* – as well as two independent gene losses on chromosomes 7 and 12
266 in both *Petunia* (Figure 6 — figure supplement 4) and in *Salpiglossis sinuata* (Figure 6 — figure
267 supplement 6) – should have happened. However, if the insertion happened after WGD, then only
268 one gene loss in *Petunia* and *Salpiglossis* needs to be invoked (Figure 6 — figure supplement 6).
269 The latter scenario is more likely based on the principle of parsimony.

270 Our ancestral state reconstruction inference supports the notion that the medium chain
271 acylsugars co-emerged with the ACS/ECH genes in the syntenic regions in the common ancestor
272 of *Solanum* (Figure 6 — figure supplement 7). This leads us to propose that the emergence of
273 both ACS and ECH genes in the synteny was a prerequisite for the rise of medium chain
274 acylsugars in Solanaceae species (Figure 6B). Consistent with the hypothesis, only short chain
275 acylsugars were observed in *Petunia* (Liu et al., 2017) (Figure 6 — figure supplement 8), which

276 correlates with the absence of ACS homolog (Figure 6B). In contrast, medium chain acylsugars
277 were detected throughout *Solanum* (Figure 6 — figure supplement 8), supported by the
278 observation that both ACS and ECH homologs are present in extant *Solanum* species (Figure 6B).
279 Interestingly, although *Nicotiana* species collectively have both ACS and ECH genes (Figure 6B),
280 they do not produce medium chain acylsugars (Figure 6 — figure supplement 8) presumably due
281 to gene losses. For example, the ECH homologs are pseudogenes in *N. benthamiana* and *N.*
282 *tomentosiformis* (Figure 6A). These results show that the presence of both functional ACS and
283 ECH genes are associated with the accumulation of medium chain acylsugars, supporting our
284 hypothesis above.

285 Although no medium chain acylsugars were detected in *Nicotiana* species examined, the
286 ACS and ECH genes may have been present in the syntenic region prior to divergence of
287 *Solanum* and *Nicotiana*. This suggests that one or more species that diverged from the common
288 ancestor of *Solanum* and *Nicotiana* could have medium chain acylsugars. To test this hypothesis,
289 we extended the phenotypic analysis to six such Solanaceae genera (Figure 6 — figure
290 supplement 8). Indeed, we found that species in *Jaltomata, Physalis, Iochroma, Atropa, and*
291 *Hyoscyamus*, which diverged from the common ancestor with *Nicotiana* but before *Solanum,*
292 have medium chain acylsugars (Figure 6 — figure supplement 8).

293 **Discussion**

294 This study identified a *S. lycopersicum* chromosome 7 synteny of ACS, ECH, and BAHD
295 acyltransferase genes including two involved in medium chain acylsugar biosynthesis. The
296 discovery of this locus was prompted by our observation of increased C10 containing acylsugars
297 in tomato recombinant lines carrying this region from the wild tomato *S. pennellii* chromosome 7.
298 *In vitro* biochemistry revealed that Sl-AACS1 produces acyl-CoAs using C6-C12 fatty acids as
299 substrates. The function of *AACS1* and *AECH1* in cultivated and wild tomato medium chain
300 acylsugar biosynthesis was confirmed by genome editing, and extended to the phylogenetically

301 distant *S. quitoense* using VIGS. The trichome tip cell-specific expression of these genes is
302 similar to that of previously characterized acylsugar pathway genes (Fan et al., 2019).
303 There are increasing examples of plant specialized metabolic innovation evolving from gene
304 duplication and neofunctionalization of primary metabolic enzymes (Maeda, 2019; Milo and Last,
305 2012; Moghe and Last, 2015; Zi et al., 2014). Recruitment of *Sl-AACS1* and *Sl-AECH1* from fatty
306 acid metabolism provides new examples of 'hijacking' primary metabolic genes into acylsugar
307 biosynthesis, in addition to an isopropylmalate synthase (*Sl-IPMS3*) and an invertase (*Sp-ASFF1*)
308 (Leong et al., 2019; Ning et al., 2015). We hypothesize that both AACS1 and AECH1 participate
309 in generation of medium chain acyl-CoAs, the acyl donor substrates for ASAT-catalyzed
310 acylsugar biosynthesis. Indeed, Sl-AACS1 exhibits *in vitro* function consistent with this
311 hypothesis, efficiently utilizing medium chain fatty acids as substrates to synthesize acyl-CoAs.
312 Strikingly, Sl-AECH1 perturbs membrane lipid composition when transiently expressed in *N.*
313 *benthamiana* leaves, generating unusual C12-chain membrane lipids.
314 These results suggest that evolution of trichome tip cell-specific gene expression potentiated
315 the co-option of *AACS1* and *AECH1* in medium chain acylsugar biosynthesis. Analogous to
316 trichomes producing medium chain acylsugars, seeds of phylogenetically diverse plants
317 accumulate medium chain fatty acid storage lipids (Ohlrogge et al., 2018). In contrast, fatty acids
318 with unusual structures, including those of medium chain lengths, are rarely found in membrane
319 lipids, presumably because these would perturb membrane bilayer structure and function (Millar
320 et al., 2000). For example, seed embryo-specific expression of three neofunctionalized enzyme
321 variant genes in *Cuphea* species – an acyl-ACP thioesterase (Dehesh et al., 1996), a 3-ketoacyl-
322 ACP synthase (Dehesh et al., 1998), and a diacylglycerol acyltransferase (Iskandarov et al., 2017)
323 – lead to production of medium chain seed storage lipids (Voelker and Kinney, 2001). Trichome
324 tip cell restricted expression of *AACS1* and *AECH1* represents an analogous example of diverting
325 neofunctionalized fatty acid enzymes from general metabolism into cell-specific specialized

326 metabolism. It is notable that we obtained evidence that Sl-AACS1 and Sl-AECH1 are targeted to
327 mitochondria. Because the other characterized acylsugar biosynthetic enzymes – ASATs and Sl-
328 IPMS3 – appear to be cytoplasmic, these results suggest that medium chain acylsugar substrates
329 are intracellularly transported within the trichome tip cell. It is worth noting that Sl-AACS1 seems
330 to show higher activity with C8 fatty acid than with C10 or C12 (Figure 4, B and C), while no C8
331 containing acylsugars were described in tomato trichomes (Ghosh et al., 2014). This suggests that
332 C8 fatty acids are not synthesized in trichomes.

333 Beyond employing functional approaches, this study demonstrates the value of a combined
334 comparative genomic and metabolomic analysis in reconstructing the evolutionary history of a
335 gene cluster: in this case over tens of millions of years. We propose that the acylsugar gene
336 cluster started with a 'founder' BAHD acyltransferase gene, followed by sequential insertion of
337 ECH and ACS (Figure 6B). This *de novo* assembly process is analogous to evolution of the
338 antimicrobial triterpenoid avenacin cluster in oat (Qi et al., 2006, 2004). There are two
339 noteworthy features of our approach. First, reconstructing acylsugar gene cluster evolution in a
340 phylogenetic context allows us to deduce cluster composition in extant species (Figure 6B).
341 Second, it links cluster genotype with acylsugar phenotype and allows inference of acylsugar
342 diversity across the Solanaceae (Figure 6 and Figure 6 — figure supplement 8).

343 The current architecture of the Solanaceae acylsugar synteny merely represents a snapshot of
344 a genomic neighborhood that is dynamic, which echoes a recent study of triterpene biosynthetic
345 gene clusters in the Brassicaceae (Liu et al., 2019; Peters, 2020). *De novo* assembly of the gene
346 cluster was accompanied by gene duplication, transposition, pseudogenization, and deletion in
347 different genera. In the case of non-acylsugar producer *Capsicum*, although phylogenetic analysis
348 revealed putative *Sl-AACS1* and *Sl-AECH1* orthologous genes, they are not harbored in the
349 syntenic region, probably due to translocation or assembly quality issues (Figure 6A and Figure 6
350 — figure supplement 1). In *Nicotiana*, the ECH genes became pseudogenized (Figure 6B), which

351 is associated with lack of detectable plant medium chain acylsugars (Figure 6 — figure
352 supplement 8). In tomatoes, the trichome expressed *Solyc07g043660* derives from a recent
353 tandem duplication (Figure 6 — figure supplement 4), yet its deletion has no effect on trichome
354 acylsugars (Figure 3 — figure supplement 1D). A parsimonious explanation is that
355 *Solyc07g043660* is experiencing functional divergence, which may eventually lead to
356 pseudogenization as observed for other genes in the syntenic region.

357 In this study, we identified an acylsugar SMGC in the context of a multiple chromosome
358 syntenic region. This synteny resulted from WGD, and the acylsugar-related genes are co-
359 expressed, and involved in the same metabolic pathway. This resembles the tomato steroidal
360 alkaloid gene cluster consisting of eight genes that are dispersed into two syntenic chromosome
361 regions (Itkin et al., 2013). In fact, this tomato alkaloid SMGC is located next to the acylsugar
362 cluster (Figure 3 — figure supplement 2), which is reminiscent of two physically adjacent
363 SMGCs in the fungus *Aspergillus* (Wiemann et al., 2013). Tomato steroidal alkaloids and
364 acylsugars both serve defensive roles in plants, but are biosynthetically and structurally distinct
365 and are stored in different tissues. This raises intriguing questions. Did the separation of acylsugar
366 and alkaloid SMGCs into two chromosomes occur contemporaneously and by the same
367 mechanism? Did this colocalization confer selective advantage through additive or synergistic
368 effects of multiple classes of defensive metabolites? Answering these questions requires
369 continued mining and functional validation of metabolic gene clusters across broader plant
370 species and analysis of impacts of clustering in evolutionary and ecological contexts.

371 **Materials and methods**

Key Resources Table				
Reagent type (species) or resource	**Designation**	**Source or reference**	**Identifiers**	**Additional information**

gene (*Solanum lycopersicum* M82)	*Sl-AACS1*	This paper	GeneBank: MT078737	Characterized and named in the results
gene (*Solanum lycopersicum* M82)	*Sl-AECH1*	This paper	GeneBank: MT078736	Characterized and named in the results
gene (*Solanum pennellii* LA0716)	*Sp-AACS1*	This paper	GeneBank: MT078735	Characterized and named in the results
gene (*Solanum pennellii* LA0716)	*Sp-AECH1*	This paper	GeneBank: MT078734	Characterized and named in the results
gene (*Solanum quitoense*)	*Sq-AACS1*	This paper	GeneBank: MT078732	Characterized and named in the results
gene (*Solanum quitoense*)	*Sq-AECH1*	This paper	GeneBank: MT078731	Characterized and named in the results
gene (*Solanum quitoense*)	*Sq_c35719*	This paper	GeneBank: MT078733	Characterized and named in the results
Software, algorithm	Trimmomatic	http://www.usadellab.org/cms/index.php?page=trimmomatic	RRID:SCR_011848	
Software, algorithm	TopHat	http://ccb.jhu.edu/software/tophat/index.shtml	RRID:SCR_013035	
Software, algorithm	Cufflinks	http://cole-trapnell-lab.github.io/cufflinks/cuffmerge/	RRID:SCR_014597	
Software, algorithm	MCScanX-transposed	http://chibba.pgml.uga.edu/mcscan2/transposed/		
Software, algorithm	RAxML	https://github.com/stamatak/standard-RAxML	RRID:SCR_006086	
Software, algorithm	Mesquite	https://www.mesquiteproject.org/	RRID:SCR_017994	

372

373 **Plant materials and trichome metabolite extraction**

374 The seeds of cultivated tomato *Solanum lycopersicum* M82 were obtained from the C.M. Rick
375 Tomato Genetic Resource Center (http://tgrc/ucdvis.edu), RRID:SCR_014954. Tomato
376 introgression lines (ILs) and tomato backcross inbred lines (BILs) were from Dr. Dani Zamir
377 (Hebrew University of Jerusalem). The tomato seeds were treated with ½ strength bleach for 30
378 minutes and washed with de-ionized water three or more times before placing on wet filter paper
379 in a Petri dish. After germination, the seedlings were transferred to peat-based propagation mix
380 (Sun Gro Horticulture) and transferred to a growth chamber for two or three weeks under 16 h
381 photoperiod, 28 °C day and 22 °C night temperatures, 50% relative humidity, and 300 μmol m^{-2} s^{-}
382 ^{1} photosynthetic photon flux density. The youngest fully developed leaf was submerged in 1 mL
383 extraction solution in a 1.5 mL screw cap tube and agitated gently for 2 min. The extraction
384 solution contains acetonitrile/isopropanol/water (3:3:2) with 0.1% formic acid and 10 μM propyl-
385 4-hydroxybenzoate as internal standard. The interactive protocol of acylsugar extraction is
386 available in Protocols.io at http://dx.doi.org/10.17504/protocols.io.xj2fkqe.

387 **DNA construct assembly and tomato transformation**

388 Assembly of the CRISPR-Cas9 constructs was as described (Brooks et al., 2014; Leong et al.,
389 2019). Two guide RNAs (gRNAs) were designed targeting one or two exons of each gene to be
390 knocked out by the CRISPR-Cas9 system. The gRNAs were synthesized gene (gBlocks) by IDT
391 (Integrated DNA Technologies, location) (Supplementary file3). For each CRISPR construct, two
392 gBlocks and four plasmids from Addgene, pICH47742::2x35S-5′UTR-hCas9 (STOP)-NOST
393 (Addgene no. 49771), pICH41780 (Addgene no. 48019), pAGM4723 (Addgene no. 48015),

394 pICSL11024 (Addgene no. 51144), were mixed for DNA assembly using the Golden Gate
395 assembly kit (NEB).

396 For *in plant*a tissue specific reporter gene analysis, 1.8 kb upstream of the annotated
397 translational start site of *Sl-AACS1* and *Sl-AECH* were amplified using the primer pairs
398 SlAACS1-pro_F/R and SlECH1-pro_F/R (Supplementary file3). The amplicon was inserted into
399 the entry vector pENTR/D-TOPO, followed by cloning into the GATEWAY vector pKGWFS7.
400 For ectopically expressing *Sp-AACS1* in the cultivated tomato CRISPR mutant *sl-aacs1*, *Sp-*
401 *AACS1* gene including 1.8 kb upstream of the translational start site of *Sp-AACS1* was amplified
402 using the primer pair SpAACS1-pro-gene_F/R (Supplementary file3). The amplicon was inserted
403 into the entry vector pENTR/D-TOPO, followed by cloning into the GATEWAY vector pK7WG.

404 The plant transformation of *S. lycopersicum* M82 and *S. pennellii* LA0716 was performed
405 using the *Agrobacterium tumefaciens* strain AGL0 following published protocols (Leong et al.,
406 2019; McCormick, 1997). The primers used for genotyping the *S. lycopersicum* M82 transgenic
407 plants harboring pK7WG or pKGWFS7 construct are listed in Supplementary file3. For
408 genotyping the *S. lycopersicum* M82 CRISPR mutants in the T1 generation, the sequencing
409 primers listed in Supplementary file3 were used to amplify the genomic regions harboring the
410 gRNAs and the resultant PCR products were sent for Sanger sequencing. For genotyping the *S.*
411 *pennellii* LA0716 CRISPR mutants in the T0 generation, the sequencing primers listed in
412 Supplementary file3 were used to amplify the genomic regions harboring the gRNAs. The
413 resulting PCR products were cloned into the pGEM-T easy vector (Promega) and transformed
414 into *E. coli*. Plasmids from at least six individual *E. coli* colonies containing the amplified
415 products were extracted and verified by Sanger sequencing.

416 **Protein subcellular targeting in tobacco mesophyll cells**

417 For protein subcellular targeting analysis, the open reading frame (ORF) of *Sl-AACS1*, *Sl-AECH1*,
418 and *Solyc07g043660* were amplified using the primers listed in Supplementary file3. These

419 amplicons were inserted into pENTR/D-TOPO respectively, followed by subcloning into the
420 GATEWAY vectors pEarleyGate102 (no. CD3-684) and pEarleyGate104 (no. CD3-686), which
421 were obtained from Arabidopsis Biological Resource Center (ABRC). For the pEarleyGate102
422 constructs, the CFP was fused to the C-terminal of the tested proteins. For the pEarleyGate104
423 constructs, the YFP was fused to the protein N-terminus. Transient expressing the tested proteins
424 was performed following an established protocol (Batoko et al., 2000) with minor modifications.
425 In brief, cultures of *A. tumefaciens* (strain GV3101) harboring the expression vectors were
426 washed and resuspended with the infiltration buffer (20 mM acetosyringone, 50 mM MES pH 5.7,
427 0.5% glucose [w/v] and 2 mM Na_3PO_4) to reach OD_{600nm} = 0.05. Four-week-old tobacco
428 (*Nicotiana tabacum* cv. Petit Havana) plants grew in 21°C and 8 h short-day conditions were
429 infiltrated, and then maintained in the same growth condition for three days before being sampled
430 for imaging. The GV3101 cultures containing the mitochondria marker MT-RFP (Nelson et al.,
431 2007) were co-infiltrated to provide the control signals for mitochondrial targeting. In separate
432 experiments, the GV3101 cultures containing the peroxisome marker RFP-PTS (Nelson et al.,
433 2007) were co-infiltrated to provide the control signals for peroxisomal targeting.

434 **Confocal Microscopy**

435 A Nikon A1Rsi laser scanning confocal microscope and Nikon NIS-Elements Advanced Research
436 software were used for image acquisition and handling. For visualizing GFP fluorescence in
437 trichomes of the tomato transformants, the excitation wavelength at 488 nm and a 505- to 525-nm
438 emission filter were used for the acquisition. For visualizing signals of fluorescence proteins in
439 the tobacco mesophyll cells, CFP, YFP and RFP, respectively, were detected by excitation lasers
440 of 443 nm, 513 nm, 561 nm and emission filters of 467-502 nm, 521-554 nm, 580-630 nm.

441 ***N. benthamiana* transient gene expression and membrane lipid analysis**

442 For *N. benthamiana* transient expression of *Sl-AACS1*, *Sl-AECH1*, and homologs of *AECH1*, the
443 ORFs of these genes were amplified using primers listed in Supplementary file3, followed by

444 subcloning into pEAQ-HT vector using the Gibson assembly kit (NEB). Linearization of pEAQ-
445 HT vector was performed by XhoI and AgeI restriction enzyme double digestion. *A. tumefaciens*
446 (strain GV3101) harboring the pEAQ-HT constructs were grown in LB medium containing 50
447 μg/mL kanamycin, 30 μg/mL gentamicin, and 50 μg/ml rifampicin at 30 °C. The cells were
448 collected by centrifugation at 5000 g for 5 min and washed once with the resuspension buffer (10
449 mM MES buffer pH 5.6, 10 mM $MgCl_2$, 150 μM acetosyringone). The cell pellet was
450 resuspended in the resuspension buffer to reach OD_{600nm} = 0.5 for each strain and was incubated
451 at room temperature for 1 h prior to infiltration. Leaves of 4 to 5-week-old *N. benthamiana* grown
452 under 16 h photoperiod were used for infiltration. Five days post infiltration, the infiltrated leaves
453 were harvested, ground in liquid nitrogen, and stored at -80 °C for later analysis.

454 The membrane lipid analysis was performed as previously described (Wang and Benning,
455 2011). In brief, the *N. benthamiana* leaf polar lipids were extracted in the organic solvent
456 containing methanol, chloroform, and formic acid (20:10:1, v/v/v), separated by thin layer
457 chromatography (TLC), converted to fatty acyl methylesters (FAMEs), and analyzed by gas-
458 liquid chromatography (GLC) coupled with flame ionization. The TLC plates (TLC Silica gel 60,
459 EMD Chemical) were activated by ammonium sulfate before being used for lipid separation.
460 Iodine vapor was applied to TLC plates after lipid separation for brief reversible staining.
461 Different lipid groups on the TLC plates were marked with a pencil and were scraped for analysis.
462 For LC/MS analysis, lipids were extracted using the buffer containing
463 acetonitrile/isopropanol/water (3:3:2) with 0.1% formic acid and 10 μM propyl-4-
464 hydroxybenzoate as the internal standard.

465 **Protein expression and ACS enzyme assay**

466 To express His-tagged recombinant protein Sl-AACS1, the full-length *Sl-AACS* ORF sequence
467 was amplified using the primer pair SlAACS1-pET28_F/R (Supplementary file3) and was cloned
468 into pET28b (EMD Millipore) using the Gibson assembly kit (NEB). The pET28b vector was

469 linearized by digesting with BamHI and XhoI to create overhangs compatible for Gibson
470 assembly. The pET28b constructs were transformed into BL21 Rosetta cells (EMD Millipore).
471 The protein expression was induced by adding 0.05 mM isopropyl *β*-D-1-thiogalactopyranoside
472 to the cultures when the OD_{600nm} = 0.5. The *E. coli* cultures were further grown overnight at 16 °C,
473 120 rpm. The His-tagged proteins were purified by Ni-affinity gravity-flow chromatography
474 using the Ni-NTA agarose (Qiagen) following the product manual.

475 Measurement of acyl-CoA synthetase activity was performed using minor modifications
476 of the coupled enzyme assay described by Schneider et al. (Schneider et al., 2005). A multimode
477 plate reader (PerkinElmer, mode EnVision 2104) compatible with the 96-well UV microplate was
478 used for the assays. The fatty acid substrates were dissolved in 5% Triton X-100 (v/v) to make 5
479 mM stock solutions. The enzyme assay premix was prepared containing 0.1 M Tris-HCl (pH 7.5),
480 2 mM dithiothreitol, 5 mM ATP, 10 mM $MgCl_2$, 0.5 mM CoA, 0.8 mM NADH, 250 μM fatty
481 acid substrate, 1 mM phosphoenolpyruvate, 20 units myokinase (Sigma-Aldrich, catalog no.
482 M3003), 10 units pyruvate kinase (Roche, 10128155001), 10 units lactate dehydrogenase (Roche,
483 10127230001), and was aliquoted 95 μL each to the 96-well microplate. The reaction was started
484 by adding 5 μL (1-2 μg) proteins. The chamber of the plate reader was set to 30 °C and the OD at
485 340 nm was recorded every 5 min for 40 min. Oxidation of NADH, which is monitored by the
486 decrease of OD_{340nm}, was calculated using the NADH extinction coefficient 6.22 cm^2 μmol^{-1}.
487 Every two moles of oxidized NADH is equivalent to one mole of acyl-CoA product generated in
488 the reaction. To measure the parameters of enzyme kinetics, the fatty acid substrate concentration
489 was varied from 0 to 500 μM, with NADH set at 1 mM. The fatty acid substrates, sodium acetate
490 (C2:0), sodium butyrate (C4:0), sodium hexanoate (C6:0), sodium octanoate (C8:0), sodium
491 decanoate (C10:0), sodium laurate (C12:0), sodium myristate (C14:0), sodium palmitate (C16:0),
492 and sodium stearate (C18:0), were purchased from Sigma-Aldrich. *Trans*-2-decenoic acid (C10:1),

493 8-methylnonanoic acid (iC10:0), 3-hydroxy lauric acid (C12:OH), and 3-hydroxy myristic acid
494 (C14:OH) were purchased from Cayman Chemical.

495 **RNA extraction, sequencing, and differential gene expression analysis**

496 Total RNA was extracted from trichomes isolated from stems and shaved stems of 7-week-old *S.*
497 *pennellii* LA0716 plants using the RNAeasy Plant Mini kit (Qiagen) and digested with DNase I.
498 A total of four RNA samples extracted from two tissues with two replicates were used for RNA
499 sequencing. The sequencing libraries were prepared using the KAPA Stranded RNA-Seq Library
500 Preparation Kit. Libraries went through quality control and quantification using a combination of
501 Qubit dsDNA high sensitivity (HS), Applied Analytical Fragment Analyzer HS DNA and Kapa
502 Illumina Library Quantification qPCR assays. The libraries were pooled and loaded onto one lane
503 of an Illumina HiSeq 4000 flow cell. Sequencing was done in a 2x150bp paired end format using
504 HiSeq 4000 SBS reagents. Base calling was done by Illumina Real Time Analysis (RTA) v2.7.6
505 and output of RTA was demultiplexed and converted to FastQ format with Illumina Bcl2fastq
506 v2.19.1.

507 The paired end reads were filtered and trimmed using Trimmomatic v0.32 (A. M. Bolger
508 et al., 2014) with the setting (LLUMINACLIP: TruSeq3-PE.fa:2:30:10 LEADING:3
509 TRAILING:3 SLIDINGWINDOW:4:30), and then mapped to the *S. pennellii* LA0716 genome
510 v2.0 (A. Bolger et al., 2014) using TopHat v1.4 (Trapnell et al., 2009) with the following
511 parameters: -p (threads) 8, -i (minimum intron length) 50, -I (maximum intron length) 5000, and -
512 g (maximum hits) 20. The FPKM (Fragments Per Kilobase of transcript per Million mapped reads)
513 values for the genes were analyzed via Cufflinks v2.2 (Trapnell et al., 2010). For differential
514 expression analysis, the HTseq package (Anders et al., 2015) in Python was used to get raw read
515 counts, then Edge R version 3.22.5 (McCarthy et al., 2012) was used to compare read counts
516 between trichome-only RNA and shaved stem RNA using a generalized linear model (glmQLFit).

517 **VIGS and qRT-PCR**

518 For VIGS analysis of *Sq-AACS1* and *Sq-AECH1* in *S. quitoense*, the fragments of these two genes,
519 as well as the phytoene desaturase (PDS) gene fragment, were amplified using the primers listed
520 in Supplementary file3, cloned into pTRV2-LIC (ABRC no. CD3-1044) using the ligation-
521 independent cloning method (Dong et al., 2007), and transformed into *A. tumefaciens* (strain
522 GV3101). The VIGS experiments were performed as described (Leong et al., 2020). In brief, the
523 Agrobacterium strains harboring pTRV2 constructs, the empty pTRV2, or pTRV1 were grown
524 overnight in separate LB cultures containing 50 µg/mL kanamycin, 10 µg/mL gentamicin, and 50
525 µg/ml rifampicin at 30 °C. The cultures were re-inoculated in the induction media (50 mM MES
526 pH5.6, 0.5% glucose [w/v], 2 mM NaH_2PO_4, 200 µM acetosyringone) for overnight growth. The
527 cells were harvested, washed, and resuspended in the buffer containing 10 mM MES, pH 5.6, 10
528 mM $MgCl_2$, and 200 µM acetosyringone with the OD_{600nm} = 1. Different cultures containing
529 pTRV2 constructs were mixed with equal volume of pTRV1 cultures prior to infiltration. The 2-
530 to 3-week-old young *S. quitoense* seedlings grown under 16 h photoperiod at 24 °C were used for
531 infiltration: the two fully expanded cotyledons were infiltrated. Approximately three weeks post
532 inoculation, the fourth true leaf of each infiltrated plant was cut in half for acylsugar
533 quantification and gene expression analysis, respectively. The onset of the albino phenotype of
534 the control group infiltrated with the PDS construct was used as a visual marker to determine the
535 harvest time and leaf selection for the experimental groups. At least fourteen plants were analyzed
536 for each construct. The trichome acylsugars were extracted using the solution containing
537 acetonitrile/isopropanol/water (3:3:2) with 0.1% formic acid and 1 µM telmisartan as internal
538 standard, following the protocol at http://dx.doi.org/10.17504/protocols.io.xj2fkqe.

539 The leaf RNA was extracted using RNeasy Plant Mini kits (Qiagen) and digested with
540 DNase I. The first-strand cDNA was synthesized by Superscript II (Thermofisher Scientific)
541 using total RNA as templates. Quantitative real-time PCR was performed to analyze the *Sq-*

542 *AACS1* or *Sq-AECH1* mRNA in *S. quitoense* leaves using the primers listed in Supplementary
543 file3. EF1α was used as a control gene. A QuantStudio 7 Flex Real-Time PCR System with Fast
544 SYBR Green Master Mix (Applied Biosystems) was used for the analysis. The relative
545 quantification method ($2^{-\Delta\Delta Ct}$) was used to evaluate the relative transcripts levels.

546 **LC/MS analysis**

547 Trichome acylsugars extracted from tomato IL and BILs were analyzed using a Shimadzu LC-
548 20AD HPLC system connected to a Waters LCT Premier ToF mass spectrometer. Ten microliter
549 samples were injected into a fused core Ascentis Express C18 column (2.1 mm × 10 cm, 2.7 µm
550 particle size; Sigma-Aldrich) for reverse-phase separation with column temperature set at 40 °C.
551 The starting condition was 90% solvent A (0.15% formic acid in water) and 10% solvent B
552 (acetonitrile) with flow rate set to 0.4 mL/min. A 7 min linear elution gradient was used: ramp to
553 40% B at 1 min, then to 100% B at 5 min, hold at 100% B to 6 min, return to 90% A at 6.01 min
554 and hold until 7 min.

555 For analyzing trichome acylsugars extracted from *S. pennellii* transgenic plants and
556 membrane lipids from *N. benthamiana*, a Shimadzu LC-20AD HPLC system connected to a
557 Waters Xevo G2-XS QToF mass spectrometer was used. The starting condition were 95% solvent
558 A (10 mM ammonium formate, pH 2.8) and 5% solvent B (acetonitrile) with flow rate set to 0.3
559 mL/min. A 7 min linear elution gradient used for acylsugar analysis was: ramp to 40% B at 1 min,
560 then to 100% B at 5 min, hold at 100% B to 6 min, return to 95% A at 6.01 min and hold until 7
561 min. A 12 min linear elution gradient used for the lipid analysis was: ramp to 40% B at 1 min,
562 then to 100% B at 5 min, hold at 100% B to 11 min, return to 95% A at 11.01 min and hold until
563 12 min.

564 For analyzing trichome acylsugars extracted from other plants, a Waters Acquity UPLC
565 was coupled to a Waters Xevo G2-XS QToF mass spectrometer. The starting condition was 95%
566 solvent A (10 mM ammonium formate, pH 2.8) and 5% solvent B (acetonitrile) with flow rate set

567 to 0.3 mL/min. A 7 min linear elution gradient was: ramp to 40% B at 1 min, then to 100% B at 5
568 min, hold at 100% B to 6 min, return to 95% A at 6.01 min and held until 7 min. A 14 min linear
569 elution gradient was: ramp to 35% B at 1 min, then to 85% B at 12 min, then to 100% B at 12.01
570 min, hold at 100% B to 13 min, return to 95% A at 13.01 min and held until 14 min.

571 For Waters LCT Premier ToF mass spectrometer, the MS settings of electrospray
572 ionization in negative mode were: 2.5 kV capillary voltage, 100°C source temperature, 350°C
573 desolvation temperature, 350 liters/h desolvation nitrogen gas flow rate, 10 V cone voltage, and
574 mass range *m/z* 50 to 1500 with spectra accumulated at 0.1 seconds/function. Three collision
575 energies (10, 40, and 80 eV) were used in separate acquisition functions to generate both
576 molecular ion adducts and fragments. For Waters Xevo G2-XS QToF mass spectrometer, the MS
577 settings of the negative ion-mode electrospray ionization were as follows: 2.00 kV capillary
578 voltage, 100 °C source temperature, 350 °C desolvation temperature, 600 liters/h desolvation
579 nitrogen gas flow rate, 35V cone voltage, mass range of *m/z* 50 to 1000 with spectra accumulated
580 at 0.1 seconds/function. Three collision energies (0, 15, and 35 eV) were used in separate
581 acquisition functions. The MS settings for positive ion-mode electrospray ionization were: 3.00
582 kV capillary voltage, 100 °C source temperature, 350 °C desolvation temperature, 600 liters/h
583 desolvation nitrogen gas flow rate, 35V cone voltage, mass range of *m/z* 50 to 1000 with spectra
584 accumulated at 0.1 seconds/function. Three collision energies (0, 15, and 45 eV) were used in
585 separate acquisition functions. The Waters QuanLynx software was used to integrate peak areas
586 of the selected ion relative to the internal standard. For quantification purpose, data collected with
587 the lowest collision energy was used in the analysis.

588 **Gene coexpression analysis**

589 The publicly available tomato RNA-seq datasets and the methods used for normalizing FPKM,
590 gene expression correlation analysis were described in a recent study (Moore et al., 2020). 926
591 RNA-seq Sequence Read Archive (SRA) files for tomato from 47 studies were downloaded from

592 National Center for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/) (Table S6
593 in (Moore et al., 2020). Reads were filtered using Trimmomatic (A. M. Bolger et al., 2014) based
594 on the sequence quality with default settings, and mapped to the tomato NCBI *S. lycopersicum*
595 genome 2.5 using TopHat (Trapnell et al., 2009). Read files with <70% mapped reads were
596 discarded. Fragments per kilobase of transcript per million mapped reads (FPKM) were calculated
597 using Cufflinks (Trapnell et al., 2010). The pipeline for FPKM calling was put in
598 https://github.com/ShiuLab/RNAseq_pipeline. The median FPKM of multiple replicates was used
599 for each sample, resulting in FPKM values in 372 samples. To draw the heatmap of gene
600 expression profiles, FPKM values of a gene across all the samples were scaled, where the
601 maximum FPKM was scaled to 1, while the minimum value was 0.

602 **Synteny scan**

603 Protein sequences of annotated genes and the corresponding annotation files in General Feature
604 Format (GFF) of 11 Solanaceae species, *Ipomoea trifida,* and *Coffea canephora* were downloaded
605 from National Center for Biotechnology Information (NCBI,
606 https://www.ncbi.nlm.nih.gov/genome/) or Solanaceae Genomics Network (SGN,
607 https://solgenomics.net/). The GFF files contain the coordinates of annotated genes on assembled
608 chromosomes or scaffolds. The sources and version numbers of sequences and GFF files used are:
609 *S. lycopersicum* ITAG3.2 (SGN) and V2.5 (NCBI), *S. pennellii* SPENNV200 (NCBI) and v2.0
610 (SGN), *S. tuberosum* V3.4 (SGN), *S. melongena* r2.5.1 (SGN), *Capsicum annuum L. zunla-1*
611 V2.0 (SGN), *C. annuum_var. glabriusculum* V2.0 (SGN), *Nicotiana attenuata* NIATTr2 (SGN),
612 *N. tomentosiformis* V01 (NCBI), *N. benthamiana* V1.0.1 (SGN), *Petunia axillaris* V1.6.2 (SGN),
613 *P. inflata* V1.0.1 (SGN), *I. trifida* V1.0 (NCBI), and *C. canephora* Vx (SGN).

614 To hypothesize the evolutionary history of genes in the acylsugar gene cluster, putative
615 pseudogenes, which are homologs to protein-coding genes but with predicted premature
616 stops/frameshifts and/or protein sequence truncation, were also identified for each species as

617 described (Wang et al., 2018). Protein sequences from *Arabidopsis thaliana*, *Oryza sativa*, and *S.*
618 *lycopersicum* were used as queries in the searches against the genomic regions of target species
619 using TBLASTN (Altschul et al., 1990). The intergenic genomic sequences were identified as
620 potential pseudogenes using the pipeline from as previously described (Campbell et al., 2014; Zou
621 et al., 2009). If one of the six-frame translated sequences of the intergenic genomic sequences had
622 significant similarity to annotated protein sequences, and had premature stops/frameshifts and/or
623 were truncated (<30% of functional paralogs), the gene was defined as a pseudogene.

624 Genome-wide syntenic analysis was conducted using annotated protein-coding genes and
625 putative pseudogenes from all the species with MCScanX-transposed (Wang et al., 2013) as
626 described (Wang et al., 2018). The MCScanX-based analysis did not lead to a syntenic block of
627 acylsugar gene cluster on chromosome 7 of *S. melongena* r2.5.1, which can be due to true absence,
628 issues with genome assembly, or lack of coverage. To verify this, protein sequences of *S.*
629 *lycopersi*cum genes in genomic blocks on chromosome 7 were searched against an updated *S.*
630 *melongena* genome from The Eggplant Genome Project (http://ddlab.dbt.univr.it/eggplant/) that
631 led to the identification of the synteny.

632 **Phylogenetic tree building**

633 Homologous genes of *Sl-AACS1* (ACS), *Sl-AECH1*(ECH), and *Solyc07g043670* (BAHD
634 acyltransferase) were obtained through BLAST (Altschul et al., 1990) search from the genomes of
635 11 Solanaceae species, *Ipomoea trifida,* and *Coffea canephora* with an Expect value threshold of
636 1e-5. To simply the phylogenetic tree, sequences which are distantly related to the target genes
637 were removed, and the remained sequences were used to rebuild the phylogenetic trees. The
638 amino acid sequences were aligned using MUSCLE (Edgar, 2004) with the default parameters.
639 The phylogenetic trees were built using the maximum likelihood method with 1000 bootstrap
640 replicates. The trees were generated using RAxML/8.0.6 (Stamatakis, 2014) with the following
641 parameters: -f a -x 12345 -p 12345 -# 1000 -m PROTGAMMAAUTO --auto-prot=bic, and were

642 shown with midpoint rooting. The final sequence alignments used to generate the phylogenetic
643 trees were provided in Supplementary file 5.

644 **Ancestral trait reconstruction**

645 Ancestral trait state reconstruction was conducted using the maximum likelihood model Mk1 in
646 Mesquite 3.6 (Massidon and Maddison, 2018). Four traits were inferred for their ancestral states.
647 They are the presence of medium chain acylsugars, presence of ACS genes in the synteny,
648 presence of ECH genes in the synteny, and presence of both ACS and ECH genes in the synteny.
649 The phylogeny of Solanaceae species was based on a previous study (Sarkinen et al., 2013).

650 **Acylsugar acyl chain composition analysis by GC-MS**

651 Acyl chains were characterized from the corresponding fatty acid ethyl esters following
652 transesterification of acylsugar extractions as previously reported (Ning et al., 2015). Plants were
653 grown for 4-8 weeks and approximately ten leaves were extracted for 3 minutes in 10 mL of 1:1
654 isopropanol:acetonitrile with 0.01% formic acid. Extractions were dried to completeness using a
655 rotary evaporator and then 300 µL of 21% (v/v) sodium ethoxide in ethanol (Sigma) was added
656 and incubated for 30 minutes with gentle rocking and vortexing every five minutes and 400 µL
657 hexane was added and vortexed for 30 seconds. To the hexane layer, 500 µL of saturated sodium
658 chloride in water was added and vortexed to facilitate a phase separation. After phase separation,
659 the top hexane layer was transferred to a new tube. The phase separation by addition of 500 µL
660 hexane was repeated twice, with the final hexane layer transferred to a 2 mL glass vial with a
661 glass insert.

662 The fatty acid ethyl esters were analyzed using an Agilent 5975 single quadrupole GC-MS
663 equipped with a 30-m, 0.25-mm internal diameter fused silica column with a 0.25-µm film
664 thickness VF5 stationary phase (Agilent). Injection of 1 µL of each hexane extract was performed
665 using splitless mode. The gas chromatography program was as follows: inlet temperature, 250°C;

666 initial column temperature, 70°C held for 2 min; ramped at 20°C/min until 270°C, then held at
667 270°C 3 min. The helium carrier gas was used with 70 eV electron ionization. Acyl chain type
668 was determined through NIST Version 2.3 library matches of the mass spectra of the
669 corresponding ethyl ester and relative abundances were determined through integrating the
670 corresponding peak area over the total acyl chain peak area.

671 **Acknowledgements**: We thank the C.M. Rick Tomato Genetics Resource Center (University of
672 California Davis, CA USA) for providing tomato seeds, Zamir lab in Hebrew University of
673 Jerusalem for providing tomato ILs and BILs seeds. We acknowledge Dr. Kun Wang and Dr.
674 Christoph Benning for their helpful guidance in lipid analysis. We thank Krystle Wiegert-
675 Rininger and Cornelius Barry for their help in RNA sequencing and Dr. Kent Chapman from
676 University of North Texas for helpful discussions. We acknowledge Kathleen Imre and Sara
677 Haller for their help with tomato transformation. We thank the MSU Center for Advanced
678 Microscopy and RTSF Mass Spectrometry and Metabolomics Core Facilities for their support
679 with LC/MS analysis.

680 **Funding**: This work was supported by the US National Science Foundation (NSF) Plant Genome
681 Research Program grant IOS-1546617 to R.L.L and S.H.S; NSF grant DEB-1655386 and the U.S.
682 Department of Energy Great Lakes Bioenergy Research Center (BER DE-SC0018409) to S.H.S;
683 NSF grant MCB1727362 and AgBioResearch to F.B. This publication was made possible by a
684 predoctoral training award to B.J.L. from Grant Number T32-GM110523 from the National
685 Institute of General Medical Sciences of the National Institutes of Health. R.C. was supported by
686 the Plant Genomics at MSU REU Program funded in part by NSF 1757043. C.A.S. was supported
687 by a postdoctoral fellowship from the NSF IOS-PGRP-1811055.

688 **Competing interests:** The authors declare that they have no competing interests.

689 **Data and materials availability:** All data needed to evaluate the conclusions in the paper are
690 present in the paper and/or the Supplementary Materials. The RNA-seq reads were deposited in
691 the National Center for Biotechnology Information Sequence Read Archive under the accession
692 number PRJNA605501. The following materials require a material transfer agreement: pEAQ-HT,
693 pK7WG, pKGWFS7, pEarleyGate102, pEarleyGate104, pTRV2-LIC, pICH47742::2x35S-
694 5'UTR-hCas9(STOP)-NOST, pICH41780, pAGM4723, and pICSL11024. Requests for
695 biological materials or data should be submitted to R.L.L. at lastr@msu.edu.
696

697 **References**

698 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool.
699 *J Mol Biol* **215**:403–410. doi:10.1016/S0022-2836(05)80360-2
700 Anders S, Pyl PT, Huber W. 2015. HTSeq — A Python framework to work with high-throughput
701 sequencing data. *Bioinformatics* **31**:166–169. doi:10.1093/bioinformatics/btu638
702 Bailey JA, Gu Z, Clark RA, Reinert K, Samonte R V., Schwartz S, Adams MD, Myers EW, Li
703 PW, Eichler EE. 2002. Recent segmental duplications in the human genome. *Science*.
704 doi:10.1126/science.1072047
705 Batoko H, Zheng HQ, Hawes C, Moore I. 2000. A Rab1 GTPase is required for transport between
706 the endoplasmic reticulum and Golgi apparatus and for normal Golgi movement in plants.
707 *Plant Cell* **12**:2201–2218. doi:10.1105/tpc.12.11.2201
708 Bolger A, Scossa F, Bolger ME, Lanz C, Maumus F, Tohge T, Quesneville H, Alseekh S,
709 Sorensen I, Lichtenstein G, Fich EA, Conte M, Keller H, Schneeberger K, Schwacke R,
710 Ofner I, Vrebalov J, Xu Y, Osorio S, Aflitos SA, Schijlen E, Jimenez-Gomez JM, Ryngajllo
711 M, Kimura S, Kumar R, Koenig D, Headland LR, Maloof JN, Sinha N, van Ham RCHJ,
712 Lankhorst RK, Mao L, Vogel A, Arsova B, Panstruga R, Fei Z, Rose JKC, Zamir D, Carrari
713 F, Giovannoni JJ, Weigel D, Usadel B, Fernie AR. 2014. The genome of the stress-tolerant

714 wild tomato species Solanum pennellii. *Nat Genet* **46**:1034–1038. doi:10.1038/ng.3046
715 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence
716 data. *Bioinformatics* **30**:2114–2120. doi:10.1093/bioinformatics/btu170
717 Boutanaev AM, Moses T, Zi J, Nelson DR, Mugford ST, Peters RJ, Osbourn A. 2015.
718 Investigation of terpene diversification across multiple sequenced plant genomes. *Proc Natl*
719 *Acad Sci USA* **112**:E81–E88. doi:10.1073/pnas.1419547112
720 Brocard C, Hartig A. 2006. Peroxisome targeting signal 1: Is it really a simple tripeptide?
721 *Biochim Biophys Acta - Mol Cell Res* **1763**:1565–1573. doi:10.1016/j.bbamcr.2006.08.022
722 Brooks C, Nekrasov V, Lippman ZB, Van Eck J. 2014. Efficient gene editing in tomato in the
723 first generation using the clustered regularly interspaced short palindromic repeats/CRISPR-
724 associated9 system. *Plant Physiol* **166**:1292–1297. doi:10.1104/pp.114.247577
725 Buchanan, B. B. Gruissem, W. Jones RL. 2015. Biochemistry and Molecular Biology of Plants,
726 2ND ed. Rockville, MD: American Society of Plant Physiologists.
727 Campbell MS, Law MY, Holt C, Stein JC, Moghe GD, Hufnagel DE, Lei J, Achawanantakun R,
728 Jiao D, Lawrence CJ, Ware D, Shiu SH, Childs KL, Sun Y, Jiang N, Yandell M. 2014.
729 MAKER-P: A Tool kit for the rapid creation, management, and quality control of plant
730 genome annotations. *Plant Physiol* **164**:513–524. doi:10.1104/pp.113.230144
731 Castillo DA, Kolesnikova MD, Matsuda SPT. 2013. An effective strategy for exploring unknown
732 metabolic pathways by genome mining. *J Am Chem Soc* **135**:5885–5894.
733 doi:10.1021/ja401535g
734 Chae L, Kim T, Nilo-Poyanco R, Rhee SY. 2014. Genomic signatures of specialized metabolism
735 in plants. *Science* **344**:510–513. doi:10.1126/science.1252076
736 D'Auria JC. 2006. Acyltransferases in plants: a good time to be BAHD. *Curr Opin Plant Biol*
737 **9**:331–340. doi:10.1016/j.pbi.2006.03.016
738 Dehesh K, Edwards P, Fillatti JA, Slabaugh M, Byrne J. 1998. KAS IV: A 3-ketoacyl-ACP

739 synthase from *Cuphea sp.* is a medium chain specific condensing enzyme. *Plant J* **15**:383–
740 390. doi:10.1046/j.1365-313X.1998.00218.x
741 Dehesh K, Jones A, Knutzon DS, Voelker TA. 1996. Production of high levels of 8:0 and 10:0
742 fatty acids in transgenic canola by overexpression of *Ch FatB2*, a thioesterase cDNA from
743 *Cuphea* hookeriana. *Plant J* **9**:167–172. doi:10.1046/j.1365-313X.1996.09020167.x
744 Dong Y, Burch-Smith TM, Liu Y, Mamillapalli P, Dinesh-Kumar SP. 2007. A ligation-
745 independent cloning tobacco rattle virus vector for high-throughput virus-induced gene
746 silencing identifies roles for *NbMADS4-1* and *-2* in floral development. *Plant Physiol*
747 **145**:1161–1170. doi:10.1104/pp.107.107391
748 Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high
749 throughput. *Nucleic Acids Res* **32**:1792–1797. doi:10.1093/nar/gkh340
750 Eshed Y, Zamir D. 1995. An introgression line population of *Lycopersicon pennellii* in the
751 cultivated tomato enables the identification and fine mapping of yield-associated QTL.
752 *Genetics* **141**:1147–1162.
753 Fan P, Leong BJ, Last RL. 2019. Tip of the trichome: evolution of acylsugar metabolic diversity
754 in Solanaceae. *Curr Opin Plant Biol* **49**:8–16. doi:10.1016/j.pbi.2019.03.005
755 Fan P, Miller AM, Schilmiller AL, Liu X, Ofner I, Jones AD, Zamir D, Last RL. 2016. In vitro
756 reconstruction and analysis of evolutionary variation of the tomato acylsucrose metabolic
757 network. *Proc Natl Acad Sci* **113**:E239–E248. doi:10.1073/pnas.1517930113
758 Frey M, Chomet P, Glawischnig E, Stettner C, Grün S, Winklmair A, Eisenreich W, Bacher A,
759 Meeley RB, Briggs SP, Simcox K, Gierl A. 1997. Analysis of a chemical plant defense
760 mechanism in grasses. *Science* **277**:696–699. doi:10.1126/science.277.5326.696
761 Ghosh B, Westbrook TC, Jones AD. 2014. Comparative structural profiling of trichome
762 specialized metabolites in tomato (*Solanum lycopersicum*) and *S. habrochaites*: acylsugar
763 profiles revealed by UHPLC/MS and NMR. *Metabolomics* **10**:496–507.

764 doi:10.1007/s11306-013-0585-y 585 [pii]
765 Herrera-Salgado Y, Va L, Rios Y, Alvarez L. 2005. *Myo*-inositol-Derived glycolipids with anti-
766 inflammatory activity from *Solanum lanceolatum*. *J Nat Prod* **68**:1031–1036.
767 Hurney SM. 2018. Strategies for profiling and discovery of acylsugar specialized metabolites.
768 Michigan State University.
769 Iskandarov U, Silva JE, Kim HJ, Andersson M, Cahoon RE, Mockaitis K, Cahoon EB. 2017. A
770 specialized diacylglycerol acyltransferase contributes to the extreme medium-chain fatty acid
771 content of *Cuphea* seed oil. *Plant Physiol* **174**:97–109. doi:10.1104/pp.16.01894
772 Itkin M, Heinig U, Tzfadia O, Bhide AJ, Shinde B, Cardenas PD, Bocobza SE, Unger T, Malitsky
773 S, Finkers R, Tikunov Y, Bovy A, Chikate Y, Singh P, Rogachev I, Beekwilder J, Giri AP,
774 Aharoni A. 2013. Biosynthesis of antinutritional alkaloids in Solanaceous crops is mediated
775 by clustered genes. *Science* **341**:175–179. doi:10.1126/science.1240230
776 Jeon JE, Kim J-G, Fischer CR, Mehta N, Dufour-Schroif C, Wemmer K, Mudgett MB, Sattely E.
777 2020. A pathogen-responsive gene cluster for highly modified fatty acids in tomato. *Cell*
778 **180**:176-187.e19. doi:10.1016/j.cell.2019.11.037
779 Khersonsky O, Tawfik DS. 2010. Enzyme promiscuity: a mechanistic and evolutionary
780 perspective. *Annu Rev Biochem* **79**:471–505. doi:10.1146/annurev-biochem-030409-143718
781 Leckie BM, D'Ambrosio DA, Chappell TM, Halitschke R, De Jong DM, Kessler A, Kennedy GG,
782 Mutschler MA. 2016. Differential and synergistic functionality of acylsugars in suppressing
783 oviposition by insect herbivores. *PLoS One* **11**:e0153345. doi:10.1371/journal.pone.0153345
784 Leong BJ, Hurney SM, Fiesel PD, Moghe GD, Jones AD, Last RL. 2020. Specialized metabolism
785 in a non-model nightshade: trichome acylinositol biosynthesis. *Plant Physiol*.
786 doi:10.1104/pp.20.00276
787 Leong BJ, Last RL. 2017. Promiscuity, impersonation and accommodation: evolution of plant
788 specialized metabolism. *Curr Opin Struct Biol* **47**:105–112. doi:10.1016/j.sbi.2017.07.005

789 Leong BJ, Lybrand DB, Lou YR, Fan P, Schilmiller AL, Last RL. 2019. Evolution of metabolic
790 novelty: A trichome-expressed invertase creates specialized metabolic diversity in wild
791 tomato. *Sci Adv* **5**:eaaw3754. doi:10.1126/sciadv.aaw3754
792 Liu X, Enright M, Barry CS, Jones AD. 2017. Profiling, isolation and structure elucidation of
793 specialized acylsucrose metabolites accumulating in trichomes of *Petunia* species.
794 *Metabolomics* **13**:1–10. doi:10.1007/s11306-017-1224-9
795 Liu Z, Suarez Duran HG, Harnvanichvech Y, Stephenson MJ, Schranz ME, Nelson D, Medema
796 MH, Osbourn A. 2019. Drivers of metabolic diversification: how dynamic genomic
797 neighbourhoods generate new biosynthetic pathways in the Brassicaceae. *New Phytol*.
798 doi:10.1111/nph.16338
799 Maeda HA. 2019. Evolutionary diversification of primary metabolism and its contribution to
800 plant chemical diversity. *Front Plant Sci* **10**:881. doi:10.3389/fpls.2019.00881
801 Mandal S, Ji W, McKnight TD. 2020. Candidate gene networks for acylsugar metabolism and
802 plant defense in wild tomato *Solanum pennellii*. *Plant Cell* **32**:81–99.
803 doi:10.1105/tpc.19.00552
804 Massidon WP, Maddison DR. 2018. Mesquite: A modular system for evolutionary analysis.
805 Version 3.4. *http://www.mesquiteproject.org*. doi:10.1017/CBO9781107415324.004
806 McCarthy DJ, Chen Y, Smyth GK. 2012. Differential expression analysis of multifactor RNA-
807 Seq experiments with respect to biological variation. *Nucleic Acids Res* **40**:4288–4297.
808 doi:10.1093/nar/gks042
809 McCormick S. 1997. Transformation of tomato with *Agrobacterium tumefaciens* In: Lindsey K,
810 editor. Plant Tissue Culture Manual. Dordrecht: Springer Netherlands. doi:10.1007/978-94-
811 009-0103-2
812 Millar AA, Smith MA, Kunst L. 2000. All fatty acids are not equal: Discrimination in plant
813 membrane lipids. *Trends Plant Sci* **5**:95–101. doi:10.1016/S1360-1385(00)01566-1

814 Milo R, Last RL. 2012. Achieving diversity in the face of constraints: Lessons from metabolism.
815 *Science* **336**:1663–1667. doi:10.1126/science.1217665
816 Mithöfer A, Boland W. 2012. Plant defense against herbivores: Chemical aspects. *Annu Rev Plant*
817 *Biol* **63**:431–450. doi:10.1146/annurev-arplant-042110-103854
818 Moghe G, Last RL. 2015. Something old, something new: Conserved enzymes and the evolution
819 of novelty in plant specialized metabolism. *Plant Physiol* **169**:1512–1523.
820 doi:10.1104/pp.15.00994
821 Moghe GD, Leong BJ, Hurney S, Jones AD, Last RL. 2017. Evolutionary routes to biochemical
822 innovation revealed by integrative analysis of a plant-defense related specialized metabolic
823 pathway. *eLife* **6**:e28468. doi:10.7554/eLife.28468
824 Moore BM, Wang P, Fan P, Lee A, Leong B, Lou Y-R, Schenck CA, Sugimoto K, Last R, Lehti-
825 Shiu MD, Barry CS, Shiu S-H. 2020. Within and cross species predictions of plant
826 specialized metabolism genes using transfer learning. *bioRxiv*.
827 doi:10.1101/2020.01.13.112102
828 Nadakuduti SS, Uebler JB, Liu X, Jones AD, Barry CS. 2017. Characterization of trichome-
829 expressed BAHD acyltransferases in *Petunia axillaris* reveals distinct acylsugar assembly
830 mechanisms within the Solanaceae. *Plant Physiol* **175**:36–50. doi:10.1104/pp.17.00538
831 Nelson BK, Cai X, Nebenführ A. 2007. A multicolored set of *in vivo* organelle markers for co-
832 localization studies in Arabidopsis and other plants. *Plant J* **51**:1126–1136.
833 doi:10.1111/j.1365-313X.2007.03212.x
834 Ning J, Moghe GD, Leong B, Kim J, Ofner I, Wang Z, Adams C, Jones AD, Zamir D, Last RL.
835 2015. A feedback insensitive isopropylmalate synthase affects acylsugar composition in
836 cultivated and wild tomato. *Plant Physiol* **160**:1821–1835. doi:10.1104/pp.15.00474
837 Noda-Garcia L, Liebermeister W, Tawfik DS. 2018. Metabolite–enzyme coevolution: From
838 single enzymes to metabolic pathways and networks. *Annu Rev Biochem* **87**:187–216.

839 doi:10.1146/annurev-biochem-062917-012023
840 Nützmann HW, Huang A, Osbourn A. 2016. Plant metabolic clusters – from genetics to genomics.
841 *New Phytol* **211**:771–789. doi:10.1111/nph.13981
842 Nützmann HW, Osbourn A. 2014. Gene clustering in plant specialized metabolism. *Curr Opin*
843 *Biotechnol* **26**:91–99. doi:10.1016/j.copbio.2013.10.009
844 Ofner I, Lashbrooke J, Pleban T, Aharoni A, Zamir D. 2016. *Solanum pennellii* backcross inbred
845 lines (BILs) link small genomic bins with tomato traits. *Plant J* **87**:151–160.
846 doi:10.1111/tpj.13194
847 Ohlrogge J, Thrower N, Mhaske V, Stymne S, Baxter M, Yang W, Liu J, Shaw K, Shorrosh B,
848 Zhang M, Wilkerson C, Matthäus B. 2018. PlantFAdb: a resource for exploring hundreds of
849 plant fatty acid structures synthesized by thousands of plants and their phylogenetic
850 relationships. *Plant J* **96**:1299–1308. doi:10.1111/tpj.14102
851 Panchy N, Lehti-Shiu MD, Shiu S-H. 2016. Evolution of gene duplication in plants. *Plant Physiol*
852 **171**:2294–2316. doi:10.1104/pp.16.00523
853 Peters RJ. 2020. Doing the gene shuffle to close synteny: dynamic assembly of biosynthetic gene
854 clusters. *New Phytol*. doi:10.1111/nph.16631
855 Pichersky E, Lewinsohn E. 2011. Convergent evolution in plant specialized metabolism. *Annu*
856 *Rev Plant Biol* **62**:549–566. doi:10.1146/annurev-arplant-042110-103814
857 Prisic S, Xu M, Wilderman PR, Peters RJ. 2004. Rice contains two disparate *ent*-copalyl
858 diphosphate synthases with distinct metabolic functions. *Plant Physiol* **136**:4228–4236.
859 doi:10.1104/pp.104.050567
860 Qi X, Bakht S, Leggett M, Maxwell C, Melton R, Osbourn A. 2004. A gene cluster for secondary
861 metabolism in oat: Implications for the evolution of metabolic diversity in plants. *Proc Natl*
862 *Acad Sci USA* **101**:8233–8238. doi:10.1073/pnas.0401301101
863 Qi X, Bakht S, Qin B, Leggett M, Hemmings A, Mellon F, Eagles J, Werck-Reichhart D, Schaller

864 H, Lesot A, Melton R, Osbourn A. 2006. A different function for a member of an ancient
865 and highly conserved cytochrome P450 family: From essential sterols to plant defense. *Proc*
866 *Natl Acad Sci USA* **103**:18848–18853. doi:10.1073/pnas.0607849103
867 Rokas A, Wisecaver JH, Lind AL. 2018. The birth, evolution and death of metabolic gene clusters
868 in fungi. *Nat Rev Microbiol* **16**:731–744. doi:10.1038/s41579-018-0075-3
869 Sainsbury F, Thuenemann EC, Lomonossoff GP. 2009. PEAQ: Versatile expression vectors for
870 easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnol J*
871 **7**:682–693. doi:10.1111/j.1467-7652.2009.00434.x
872 Sarkinen T, Bohs L, Olmstead RG, Knapp S, Särkinen T, Bohs L, Olmstead RG, Knapp S. 2013.
873 A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated
874 1000-tip tree. *BMC Evol Biol* **13**:214. doi:10.1186/1471-2148-13-214
875 Schenck CA, Last RL. 2019. Location, location! cellular relocalization primes specialized
876 metabolic diversification. *FEBS J*. doi:10.1111/febs.15097
877 Schilmiller A, Shi F, Kim J, Charbonneau AL, Holmes D, Daniel Jones A, Last RL. 2010. Mass
878 spectrometry screening reveals widespread diversity in trichome specialized metabolites of
879 tomato chromosomal substitution lines. *Plant J* **62**:391–403. doi:10.1111/j.1365-
880 313X.2010.04154.x
881 Schilmiller AL, Charbonneau AL, Last RL. 2012. Identification of a BAHD acetyltransferase that
882 produces protective acyl sugars in tomato trichomes. *Proc Natl Acad Sci USA* **109**:16377–82.
883 doi:10.1073/pnas.1207906109
884 Schilmiller AL, Moghe GD, Fan P, Ghosh B, Ning J, Jones AD, Last RL. 2015. Functionally
885 divergent alleles and duplicated loci encoding an acyltransferase contribute to acylsugar
886 metabolite diversity in *Solanum* trichomes. *Plant Cell* **27**:1002–1017.
887 doi:10.1105/tpc.15.00087
888 Schläpfer P, Zhang P, Wang C, Kim T, Banf M, Chae L, Dreher K, Chavali AK, Nilo-Poyanco R,

889 Bernard T, Kahn D, Rhee SY. 2017. Genome-wide prediction of metabolic enzymes,
890 pathways, and gene clusters in plants. *Plant Physiol* **173**:2071–2059.
891 doi:10.1104/pp.16.01942
892 Schneider K, Kienow L, Schmelzer E, Colby T, Bartsch M, Miersch O, Wasternack C, Kombrink
893 E, Stuible HP. 2005. A new type of peroxisomal acyl-coenzyme a synthetase from
894 Arabidopsis thaliana has the catalytic capacity to activate biosynthetic precursors of
895 jasmonic acid. *J Biol Chem* **280**:13962–13972. doi:10.1074/jbc.M413578200
896 Schneider LM, Adamski NM, Christensen CE, Stuart DB, Vautrin S, Hansson M, Uauy C, Von
897 Wettstein-Knowles P. 2016. The *Cer-cqu* gene cluster determines three key players in a β-
898 diketone synthase polyketide pathway synthesizing aliphatics in epicuticular waxes. *J Exp*
899 *Bot* **67**:2715–2730. doi:10.1093/jxb/erw105
900 Schuurink R, Tissier A. 2019. Glandular trichomes: micro-organs with model status? *New Phytol*
901 **225**:2251–2266. doi:10.1111/nph.16283
902 Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of
903 large phylogenies. *Bioinformatics* **30**:1312–1313. doi:10.1093/bioinformatics/btu033
904 Takos AM, Knudsen C, Lai D, Kannangara R, Mikkelsen L, Motawia MS, Olsen CE, Sato S,
905 Tabata S, Jørgensen K, Møller BL, Rook F. 2011. Genomic clustering of cyanogenic
906 glucoside biosynthetic genes aids their identification in *Lotus japonicus* and suggests the
907 repeated evolution of this chemical defence pathway. *Plant J* **68**:273–286.
908 doi:10.1111/j.1365-313X.2011.04685.x
909 Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: Discovering splice junctions with RNA-Seq.
910 *Bioinformatics* **25**:1105–1111. doi:10.1093/bioinformatics/btp120
911 Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, Van Baren MJ, Salzberg SL, Wold BJ,
912 Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated
913 transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**:511–515.

914 doi:10.1038/nbt.1621

915 Voelker T, Kinney AJ. 2001. Variations in the biosynthesis of seed-storage lipids. *Annu Rev Plant*
916 *Physiol Plant Mol Biol* **52**:335–361. doi:10.1146/annurev.arplant.52.1.335

917 Wang P, Moore BM, Panchy NL, Meng F, Lehti-Shiu MD, Shiu SH. 2018. Factors influencing
918 gene family size variation among related species in a plant family, Solanaceae. *Genome Biol*
919 *Evol* **10**:2596–2613. doi:10.1093/gbe/evy193

920 Wang Y, Li J, Paterson AH. 2013. MCScanX-transposed: Detecting transposed gene duplications
921 based on multiple colinearity scans. *Bioinformatics* **29**:1458–1460.
922 doi:10.1093/bioinformatics/btt150

923 Wang Z, Benning C. 2011. *Arabidopsis thaliana* polar glycerolipid profiling by thin layer
924 chromatography (TLC) coupled with gas-liquid chromatography (GLC). *J Vis Exp* 2–7.
925 doi:10.3791/2518

926 Weng J-K, Philippe RN, Noel JP. 2012. The rise of chemodiversity in plants. *Science* **336**:1667–
927 1670. doi:10.1126/science.1217411

928 Weng JK. 2014. The evolutionary paths towards complexity: A metabolic perspective. *New*
929 *Phytol* **201**:1141–1149. doi:10.1111/nph.12416

930 Wiemann P, Guo CJ, Palmer JM, Sekonyela R, Wang CCC, Keller NP. 2013. Prototype of an
931 intertwined secondary-metabolite supercluster. *Proc Natl Acad Sci USA* **110**:17065–17070.
932 doi:10.1073/pnas.1313258110

933 Wilderman PR, Xu M, Jin Y, Coates RM, Peters RJ. 2004. Identification of *syn*-pimara-7,15-
934 diene synthase reveals functional clustering of terpene synthases involved in rice
935 phytoalexin/allelochemical biosynthesis. *Plant Physiol* **135**:2098–2105.
936 doi:10.1104/pp.104.045971

937 Winzer T, Gazda V, He Z, Kaminski F, Kern M, Larson TR, Li Y, Meade F, Teodor R, Vaistij FE,
938 Walker C, Bowser TA, Graham IA. 2012. A papaver somniferum 10-gene cluster for

939 synthesis of the anticancer alkaloid noscapine. *Science* **6089**:1704–1708.
940 doi:10.1126/science.1220757
941 Wisecaver JH, Borowsky AT, Tzin V, Jander G, Kliebenstein DJ, Rokas A. 2017. A global
942 coexpression network approach for connecting genes to specialized metabolic pathways in
943 plants. *Plant Cell* **29**:944–959. doi:10.1105/tpc.17.00009
944 Zi J, Mafu S, Peters RJ. 2014. To gibberellins and beyond! Surveying the evolution of
945 (di)terpenoid metabolism. *Annu Rev Plant Biol* **65**:259–286. doi:10.1146/annurev-arplant-
946 050213-035705
947 Zou C, Lehti-Shiu MD, Thibaud-Nissen F, Prakash T, Buell CR, Shiu SH. 2009. Evolutionary
948 and expression signatures of pseudogenes in Arabidopsis and rice. *Plant Physiol* **151**:3–15.
949 doi:10.1104/pp.109.140632
950
951
952 ## Main Figure Legends
953 **Figure 1. Primary metabolites are biosynthetic precursors of tomato trichome acylsugars.** In
954 cultivated tomatoes, the trichome acylsucroses are synthesized by four Sl-ASATs using the primary
955 metabolites – sucrose and different types of acyl-CoAs – as substrates. In this study we provide evidence
956 that medium chain fatty acids are converted to acyl-CoAs by an acyl-CoA synthetase for medium chain
957 acylsugar biosynthesis.

958 **Figure 2. Mapping of a genetic locus related to acylsugar variations in tomato interspecific**
959 **introgression lines.** (**A**) Electrospray ionization negative (ESI^-) mode, base-peak intensity (BPI) LC/MS
960 chromatogram of trichome metabolites from cultivated tomato *S. lycopersicum* M82 and introgression line
961 IL7-4. The orange bars highlight two acylsugars that have higher abundance in IL7-4 than in M82. For the
962 acylsucrose nomenclature, “S” refers to a sucrose backbone, “3:22” means three acyl chains with twenty-
963 two carbons in total. The length of each acyl chain is shown in the parentheses. (**B**) Peak area percentage
964 of seven major trichome acylsugars in M82 and IL7-4. The sum of the peak area percentage of each
965 acylsugar is equal to 100% in each sample. The data is shown for three plants ± SEM. $^{**}p < 0.01$, Welch
966 two-sample *t* test. ***Figure 2 — source data 1*** includes values for the analysis. (**C**) Mapping the genetic

967 locus contributing to the IL7-4 acylsugar phenotype using selected backcross inbred lines (BILs) that have
968 recombination break points within the introgression region of IL7-4. (**D**) Narrowing down candidate genes
969 in the locus using trichome/stem RNA-seq datasets generated from previous study (Ning et al., 2015). A
970 region with duplicated genes of three types – acyl-CoA synthetase (ACS), BAHD acyltransferase, and
971 enoyl-CoA hydratase (ECH) – is shown. The red-blue color gradient provides a visual marker to rank the
972 expression levels represented by Fragments Per Kilobase of transcript per Million mapped reads (FPKM).
973 Coexpression analysis of tomato ACS, ECH, and BAHD acyltransferase family genes is shown in ***Figure 2***
974 ***— figure supplement 1***.

975 **Figure 3. CRISPR/Cas9-mediated gene knockout of tomato *Sl-AACS1* or *Sl-AECH1* eliminates**
976 **detectable medium chain containing acylsugars.** (**A**) Combined LC/MS extracted ion chromatograms
977 of trichome metabolites from CRISPR mutants *sl-aacs1* and *sl-aech1.* The medium chain acylsugars that
978 are not detected in the two mutants are denoted by pairs of vertical dotted lines. ***Figure 3 — figure***
979 ***supplement 1*** describes the design of the gRNAs and details of the gene edits. (**B**) Quantification of
980 seven major trichome acylsugars in *sl-aacs1* and *sl-aech1* mutants. Two independent T2 generation
981 transgenic lines for each mutant were used for analysis. The peak area/internal standard (IS) normalized
982 by leaf dry weight (DW) is shown from six plants ± SEM. ***Figure 3 — source data 1*** includes values for the
983 analysis. (**C**) Confocal fluorescence images showing that GFP fluorescence driven by *Sl-AACS1* or *Sl-*
984 *AECH1* is located in the tip cells of type I/IV trichomes. Their tissue specific expressions are similar to *Sl-*
985 *ASAT1* (Fan et al., 2016), which locates in a chromosome 12 region that is syntenic to the locus containing
986 *Sl-AACS1* and *Sl-AECH1*. ***Figure 3 — figure supplement 2*** provides the detailed information of the
987 syntenic region. *Sl-AACS1*, *Sl-AECH1*, and *Sl-ASAT1* are the only gene models with demonstrated
988 functions in acylsugar biosynthesis.

989 **Figure 4. Functional analysis of Sl-AACS1 and Sl-AECH1 in *N. benthamiana* and recombinant Sl-**
990 **AACS1 enzyme analysis.** (**A**) Confocal images of co-expression analysis in tobacco leaf epidermal cells
991 using C-terminal CFP-tagged either Sl-AACS1 or Sl-AECH1 and the mitochondrial marker MT-RFP.
992 Arrowheads point to mitochondria that are indicated by MT-RFP fluorescent signals. Scale bar equals 10
993 μm. ***Figure 4 — figure supplement 1B*** describes that the expressed YFP-recombinant proteins were not
994 co-localized with the peroxisomal marker RFP-PTS (**B**) Aliphatic fatty acids of different chain lengths were
995 used as the substrates to test Sl-AACS1 acyl-CoA synthase activity. Mean amount of acyl-CoAs generated

996 (nmol min^{-1} mg^{-1} proteins) was used to represent enzyme activities. The results are from three
997 measurements ± SEM. ***Figure 4 — source data 1*** includes values for the measurements. (**C**) Enzyme
998 activity of Sl-AACS1 for six fatty acid substrates. (**D**) Identification of membrane lipid phosphatidylcholine
999 (PC), which contains medium acyl chains, following transient expression of *Sl-AECH1* in *N. benthamiana*
1000 leaves. The results from expressing *Sl-AACS1* and co-expressing both *Sl-AECH1* and *Sl-AACS1* are also
1001 shown. Mole percentage (Mol %) of the acyl chains from membrane lipids with carbon number 12, 14, 16,
1002 and 18 are shown for three biological replicates ± SEM. *$p < 0.05$, **$p < 0.01$. Welch two-sample *t* test
1003 was performed comparing with the empty vector control. ***Figure 4 — source data 2*** includes values for
1004 the lipid analysis. Acyl groups of the same chain lengths with saturated and unsaturated bonds were
1005 combined in the calculation. ***Figure 4 — figure supplement 2*** shows that the putative *Sl-AECH1*
1006 orthologs from *S. pennellii* and *S. quitoense* generated medium chain lipids in the infiltrated leaves.

1007 **Figure 5. *AACS1* and *AECH1* are evolutionarily conserved in *Solanum* plants.** (**A**) A conserved
1008 syntenic genomic region containing *AACS1* and *AECH1* was found in three selected *Solanum* species.
1009 Nodes representing estimated dates since the last common ancestors (Sarkinen et al., 2013) shown on
1010 the left. The closest homologs of *AACS1* and *AECH1* in *Solanum quitoense* are shown without genomic
1011 context because the genes were identified from RNA-seq and genome sequences are not available. The
1012 lines connect genes representing putative orthologs across the four species. The trichome/stem RNA-seq
1013 data of two biological *S. pennellii* replicates are summarized (***Supplementary file2***) for genes in the
1014 syntenic region. The red-blue color gradient provides a visual marker to rank the expression levels in
1015 FPKM. Structures of representative medium chain acylsugars from *S. quitoense* (acylinositol, I4:26)
1016 (Hurney, 2018) and *S. pennellii* (acylglucose, G3:19) (Leong et al., 2019) are on the right. ***Figure 5 —***
1017 ***figure supplement 1*** shows that stable *Sp-AACS1* transformation of the M82 CRISPR mutant *sl-aacs1*
1018 restores C12 containing acylsugars (**B**) CRISPR/Cas9-mediated gene knockout of *Sp-AACS1* or *Sp-*
1019 *AECH1* in *S. pennellii* produce no detectable medium chain containing acylsugars. The ESI^+ mode LC/MS
1020 extracted ion chromatograms of trichome metabolites are shown for each mutant. The *m/z* 127.01 (left
1021 panel) corresponds to the glucopyranose ring fragment that both acylsucroses and acylglucoses generate
1022 under high collision energy positive-ion mode. The *m/z* 155.14 (center panel) and 183.17 (right panel)
1023 correspond to the acylium ions from acylsugars with chain length of C10 and C12, respectively. ***Figure 5***
1024 ***— figure supplement 2A-C*** describes the design of the gRNAs and the detailed information of gene edits.
1025 (**C**) Silencing *Sq-AACS1* (*Sq-c34025*) or *Sq-AECH1* (*Sq-c37194*) in *S. quitoense* using VIGS leads to

1026 reduction of total acylsugars. The peak area/internal standard (IS) normalized by leaf dry weight was
1027 shown from sixteen plants ± SEM. ***p < 0.001, Welch two-sample *t* test. ***Figure 5 — figure supplement***
1028 ***2E and 2F*** describes the VIGS experimental design and the representative LC/MS extracted ion
1029 chromatograms of *S. quitoense* major acylsugars. (**D**) Reduced gene expression of *Sq-AACS1* or *Sq-*
1030 *AECH1* correlates with decreased acylsugar levels in *S. quitoense*. The qRT-PCR gene expression data
1031 are plotted with acylsugar levels of the same leaf as described in ***Figure 5 — figure supplement 2E***.
1032 ***Figure 5 — source data 1*** includes raw data for the *S. quitoense* VIGS experiments.

1033 **Figure 6. Evolution of the acylsugar gene cluster is associated with acylsugar acyl chain diversity**
1034 **across the Solanaceae family.** (**A**) The acylsugar gene cluster syntenic regions of 11 Solanaceae
1035 species and two outgroup species *Ipomea trifida* (Convolvulaceae) and *Coffea canephora* (Rubiaceae).
1036 This is a simplified version adapted from ***Figure 6 — figure supplement 1***. Only genes from the three
1037 families – ACS (blue), BAHD acyltransferase (green), and ECH (orange) – are shown. For information
1038 about the syntenic region size in each species refer to ***Figure 6 — figure supplement 1*** and
1039 ***Supplementary file 4***. (**B**) The evolutionary history of the acylsugar gene cluster and its relation to the
1040 acylsugar phenotypic diversity. The evolution of BAHD acyltransferase genes is inferred based on ***Figure***
1041 ***6 — figure supplement 2*** and ***Figure 6 — figure supplement 5***. ECH genes based on ***Figure 6 —***
1042 ***figure supplement 3***. ACS genes based on ***Figure 6 — figure supplement* 4** and ***Figure 6 — figure***
1043 ***supplement 6***. The temporal order for the emergence for the three types of genes are shown in the
1044 colored boxes on the left: green box (BAHD acyltransferase), orange box (ECH), blue box (ACS). The
1045 yellow star represents the Solanaceae-specific WGD. Structures of representative short and medium chain
1046 acylsugars were shown on the right. ***Figure 6 — figure supplement 8*** describes distribution of acylsugar
1047 acyl chains with different lengths in species across the Solanaceae family.

1048 Legends for supplementary figures and files

1049 **Figure 2 — figure supplement 1. Expression profiles of tomato ACS, ECH, and BAHD**
1050 **acyltransferase family genes used for phylogenetic analysis in this study.** Each column represents
1051 one transcriptomic profiling dataset generated by RNA-seq analysis using samples from the cultivated
1052 tomato (*S. lycopersicum*). A total of 372 RNA-seq datasets were used for the analysis as described in the
1053 previous study (Moore et al., 2020). The normalized FPKM value of each gene across all the 372 datasets
1054 was illustrated by the color scale. The maximum FPKM value was set to 1 (red), while the minimum FPKM

1055 value was 0 (blue). The datasets generated using tomato root hairs (first column) or trichomes (second
1056 column) were indicated by arrows. Genes (y-axis) and expression datasets (x-axis) were both grouped
1057 using hierarchical clustering. Genes involved in acylsugar biosynthesis, such as *Sl-ASATs*, *Sl-AACS1*, and
1058 *Sl-AECH1*, were clustered together, which are highlighted by the orange box. Hierarchical clustering also
1059 revealed another group of genes that are root hair-specific as pointed out by the purple box. The gene ID
1060 was colored based on which gene family it belongs. Blue: ACS genes; red: ECH genes; green: BAHD
1061 acyltransferase genes.

1062 **Figure 3 — figure supplement 1. CRISPR-Cas9-mediated gene knockouts in cultivated tomato *S.*
1063 *lycopersicum*.** The gRNAs targeting *Sl-AACS1* (**A**), *Sl-AECH1*(**B**), and *Solyc07g043660* (**C**) in cultivated
1064 tomato are highlighted with red lines and text. Two pairs of gRNAs were designed that target *Sl-AECH1*,
1065 which were followed by two trials of plant transformation. DNA sequences of the self-crossed T1
1066 generation transgenic lines carrying homozygous gene edits are shown beneath the gene model. Dotted
1067 rectangle boxes highlight edited sequences. (**D**) Electrospray ionization negative (ESI^-) mode, LC/MS
1068 extracted ion chromatograms of seven major trichome acylsugars of the CRISPR mutant *solyc07g043660*
1069 and the M82 parent.

1070 **Figure 3 — figure supplement 2. The syntenic region of cultivated tomato *S. lycopersicum***
1071 **chromosome 7 and 12 harboring the acylsugar and steroidal glycoalkaloid gene clusters.** The gene
1072 models are represented by rectangles, with pseudogenes labeled with dotted lines. The lines linking the
1073 gene models of the two chromosomes denote putative orthologous genes in the synteny. GAME genes
1074 involved in glycoalkaloid metabolism were previously reported (Itkin et al., 2013).

1075 **Figure 4 — figure supplement 1. Characterization of cluster genes using leaf transient expression:**
1076 **protein subcellular targeting and impacts on lipid metabolism.** (**A**) Confocal images of co-expression
1077 analysis in *N. tabacum* leaf epidermal cells using C-terminal CFP-tagged Solyc07g043660 and the
1078 mitochondrial marker MT-RFP. Arrowheads point to mitochondria that are indicated by MT-RFP
1079 fluorescent signals. White bar equals 10 μm. (**B**) Confocal images of co-expression analysis in *N. tabacum*
1080 leaf epidermal cells using N-terminal YFP-tagged Sl-AACS1, Sl-AECH1, or Solyc07g043660, and the
1081 peroxisomal marker RFP-PTS. Arrowheads point to peroxisomes that are indicated by RFP-PTS
1082 fluorescent signals. Bar equals 10 μm. (**C**) *N. benthamiana* leaf membrane lipid acyl chain composition.
1083 The results from infiltrating *Sl-AACS1* or *Sl-AECH1* individually, and infiltrating *Sl-AECH1* and *Sl-AACS1*

1084 together are shown. The lipid abbreviations are: phosphatidylglycerol (PG), sulfoquinovosyl diacylglycerol
1085 (SQDG), digalactosyldiacylglycerol (DGDG), monogalactosyldiacylglycerol (MGDG), phosphtatidylinositols
1086 (PI), and phosphtatidylethanolamine (PE). Mole percentage (Mol %) of the acyl chains from membrane
1087 lipids with carbon number 12, 14, 16, and 18 are shown for three biological replicates ± SE. *p < 0.05, **p
1088 < 0.01. Welch two-sample *t* test was performed comparing each experimental with the empty vector
1089 control. PI and PE lipids were adjacent on the TLC plates and were pooled for analysis. Acyl groups of the
1090 same chain lengths with saturated and unsaturated bonds were combined in the calculation.

1091 **Figure 4 — figure supplement 2. The closest homologs of *Sl-AECH1* from other *Solanum* species**
1092 **generate medium chain lipids when transiently expressed in *N. benthamiana*.** (**A**) ESI$^-$ mode, LC/MS
1093 extracted ion chromatograms of two SQDGs with C12 as peaks diagnostic of lipids containing medium
1094 chain fatty acids. (**B**) Mass spectra of two SQDGs contain C12 acyl chain. Fragmentation of SQDG (16:0,
1095 12:0) and SQDG (18:3, 12:0) in ESI$^-$ mode revealed the fragment ion C12 fatty acid (*m/z*: 199.17) and the
1096 SQDG head group (*m/z*: 225.0). (**C**) Among the *Sl-AECH1* homologs tested, *Sopen07g023250* (*Sp-*
1097 *AECH1*) and *Sq_c37194* (*Sq-AECH1*), from *S. pennellii* and *S. quitoense* respectively, generated medium
1098 chain SQDG in the infiltrated leaves. The close homologs of *Sl-AECH1* in *S. lycopersicum*, *S. pennellii*,
1099 and *S. quitoense*, which also have trichome expressions, were used to build the phylogenetic tree. The
1100 nucleotide sequences were aligned with MEGA7 (www.megasoftware.net) using the default MUSCLE
1101 algorithm. The T92+G maximum likelihood model was selected for phylogenetic tree construction from 24
1102 different nucleotide substitution models based on the lowest Bayesian Information Criterion. The bootstrap
1103 values were obtained with 1000 replicates. The closest *Sl-AECH1* Arabidopsis homolog *AT1G06550*
1104 serves as an outgroup.

1105 **Figure 5 — figure supplement 1. Stable *Sp-AACS1* transformation of the M82 CRISPR mutant *sl-***
1106 ***aacs1* restores C12 containing acylsugars.** (**A**) ESI$^-$ mode, LC/MS extracted ion chromatograms are
1107 shown for seven major acylsugar peaks extracted from trichome of *S. lycopersicum* M82, *sl-aacs1*, and
1108 two independent T0 generation suppressed *sl-aacs1* transgenic lines expressing *Sp-AACS1* under its own
1109 promoter. (**B**) Peak area percentage of seven major trichome acylsugars of the suppression transgenic
1110 plants. The sum of the peak area percentage of each acylsugar equals to 100%. The results of acylsugar
1111 peak area percentage were calculated from six independent T0 transgenic lines ± SE.

1112 **Figure 5 — figure supplement 2. Functional analysis of *AACS1* and *AECH1* in *S. pennellii* and *S.***
1113 ***quitoense* via CRISPR-Cas9 system and VIGS, respectively.** The design of gRNAs targeting wild
1114 tomato *S. pennellii* LA0716 *Sp-AACS1* (**A**), *Sopen07g023250* (**B**), and *Sp-AECH1* (**C**) is highlighted with
1115 red lines and text. The transgenic T0 generation carrying chimeric or biallelic gene edits are shown
1116 beneath the gene model. DNA sequence of the gene edits was obtained through Sanger sequencing of
1117 cloned plant DNA fragments. The gene edits are highlighted with dotted rectangular boxes. (**D**) ESI^+ mode,
1118 LC/MS extracted ion chromatograms shown for C10 (*m/z*: 155.14) and C12 (*m/z*: 183.17) fatty acid ions
1119 corresponding to medium chain trichome acylsugars extracted from the CRISPR mutants
1120 *sopen07g023220* and the *S. pennellii* LA0716 parent. (**E**) Experimental design of VIGS in *S. quitoense*. A
1121 control group silencing the PDS genes was performed in parallel with the experimental groups. The onset
1122 of the albino phenotype of the control group was used as a visual marker to determine the harvest time
1123 and leaf selection in the experimental groups. The fourth true leaves were harvested and cut in half for
1124 gene expression analysis and acylsugar quantification, respectively. (**F**) ESI^- mode, LC/MS extracted ion
1125 chromatograms of six major acylsugars of *S. quitoense* for the experimental group. The three LC/MS
1126 chromatograms show representative acylsugar profiles of the empty vector control plants and the VIGS
1127 plants targeting *Sq-AACS1* and *Sq-AECH1*.

1128 **Figure 6 — figure supplement 1. Syntenic regions containing the acylsugar gene cluster.** The
1129 species name and chromosome/scaffold identifier are indicated with *S. lycopersinum* in blue font.
1130 Rectangle: protein-coding gene (solid line) or pseudogene (dotted line) colored according to the type of
1131 genes. Line connecting two genes: putative orthologous genes. Numbers underneath chromosomes:
1132 chromosome coordinates in million bases (Mb). The gene ID and location information used to generate the
1133 synteny figure is provided in **Supplementary file 4**.

1134 **Figure 6 — figure supplement 2. Analysis of the evolutionary history of the BAHD acyltransferases**
1135 **in the syntenic regions in different Solanaceae species.** (**A**) Phylogenetic tree of BAHD
1136 acyltransferases homologous to *Solyc07g043670.* Genes colored with green and labeled with green
1137 rectangles are from the syntenic regions of Solanaceae species shown in panel (B). Genes marked with
1138 stars have been biochemically tested involved in acylsugar biosynthesis in previous studies. (**B**) The
1139 acylsugar gene cluster syntenic regions of 11 Solanaceae species and two outgroup species *Ipomea*
1140 *trifida* (Convolvulaceae) and *Coffea canephora* (Rubiaceae). (**C**) Reconciled evolutionary history of BAHD

1141 acyltransferases based on panel (A) and (B). The colors of the branches correspond to different lineages
1142 shown in panel (A). Grey branch means enzymes from that lineage could not be found through BLAST and
1143 may have been lost. Numbers in the circle indicate the nodes in the phylogenetic tree as shown in panel A.
1144 The inferred evolutionary events were shown next to the nodes. Grey box highlighted the evolutionary
1145 history of the genes in the syntenic region. Before the Solanaceae specific whole genome duplication
1146 (WGD) events, the BAHD acyltransferase gene was tandemly duplicated. The WGD events resulted in at
1147 least two genomic regions (Chr07 and Chr12), each containing two BAHD acyltransferase genes. Before
1148 the divergence of *Solanum*, *Nicotiana*, and *Petunia* species, one of the tandem copies in Chr12 region was
1149 lost, and only orthologs of *Sl-ASAT1* was retained. The question mark next to *Petunia* denotes the
1150 inconsistence of the phylogenetic relationship and the chromosome location of the gene *Pa-B816* with an
1151 unknow mechanism.

1152 **Figure 6 — figure supplement 3. Analysis of the evolutionary history of the ECH genes in the**
1153 **syntenic regions in different Solanaceae species.** (**A**) Phylogenetic tree of ECH genes homologous to
1154 *Sl-AECH1*. Genes colored with orange and labeled with orange rectangles are from the syntenic regions of
1155 Solanaceae species shown in panel (B). Genes marked with stars have been tested involved in acylsugar
1156 biosynthesis. (**B**) The acylsugar gene cluster syntenic regions of 11 Solanaceae species and two outgroup
1157 species *Ipomea trifida* (Convolvulaceae) and *Coffea canephora* (Rubiaceae). (**C**) Reconciled evolutionary
1158 history of ECH enzyme based on panel (A) and (B). The colors of the branches correspond to different
1159 lineages shown in panel A. Grey branch means enzymes from that lineage could not be found through
1160 BLAST and may have been lost. Numbers in the circle indicate the nodes in the phylogenetic tree as
1161 shown in panel (A). The inferred evolutionary events were shown next to the nodes. Grey box highlighted
1162 the evolutionary history of the genes in the syntenic regions. Before the Solanaceae specific WGD events,
1163 an ECH was inserted into the syntenic region through unknown mechanism. After the WGD events, there
1164 was one ECH gene in each syntenic region on Chr07 and Chr12. Before the divergence of *Solanum*,
1165 *Nicotiana*, and *Petunia* species, the ECH gene on Chr07 had experienced a tandem duplication event,
1166 leading to two branches on the phylogenetic tree. During the speciation, the ECH gene on Chr12 was
1167 deleted from the genome in the most recent common ancestor of *Nicotiana* and *Solanum* after divergent
1168 from *Petunia*, while one of the tandem duplicates on Chr07 was lost in *Petunia*.

1169 **Figure 6 — figure supplement 4. Analysis of the evolutionary history of the ACS genes in the**
1170 **syntenic regions in different Solanaceae species.** (**A**) Phylogenetic tree of ACS genes homologous to
1171 *Sl-AACS1*. Genes colored with blue and labeled with blue rectangles are from the syntenic regions of
1172 Solanaceae species shown in panel (B). Genes marked with stars have been tested involved in acylsugar
1173 biosynthesis. (**B**) The acylsugar gene cluster syntenic regions of 11 Solanaceae species and two outgroup
1174 species *Ipomea trifida* (Convolvulaceae) and *Coffea canephora* (Rubiaceae). (**C**) Reconciled evolutionary
1175 history of ACS enzyme based on panel (A) and (B). The colors of the branches correspond to different
1176 lineages shown in panel (A). Grey branch means enzymes from that lineage could not be found through
1177 BLAST and may have been lost. Numbers in the circle indicate the nodes in the phylogenetic tree as
1178 shown in panel A. The inferred evolutionary events were shown next to the nodes. A tandem duplication
1179 event happened before the Solanaceae specific WGD events, leading to two adjacent ACS genes on
1180 Chr02 (*Solyc02g082880* and *Solyc02g082870*), which were placed on two independent lineages in the
1181 phylogenetic tree. *Solyc02g082870* had gone through two rounds of WGD events, supported by the
1182 observation that *Solyc02g082870* and *Solyc03g032210* are located in corresponding syntenic blocks.
1183 *Solyc02g082880* may have experienced the segmental duplication, resulting in the ACS gene on Chr07,
1184 which had experienced another two rounds of tandem duplication in the common ancestor of *Solanum*
1185 species (*Sl-AACS1*, *Solyc07g043660*, and *Solyc07g043640*). However, whether the segmental duplication
1186 event happened before or after the Solanaceae specific WGD events cannot be well resolved by the
1187 phylogenetic analysis. Two hypotheses were proposed as shown in the grey boxes. If the insertion
1188 happened before WGD, two independent gene loss events on chromosomes 7 and 12 should have
1189 happened in Petunia (Hypothesis 2). If the insertion happened after WGD, only one gene loss in Petunia
1190 was supposed to have happened (Hypothesis 1). Note that node 4 in (A) leads to two clades, one without
1191 any *Petunia* ACS homolog (darker blue) and the other with *Petunia* homologs (cyan). With regard to the
1192 timing of the duplication event leading to these two clades, it was likely before the split between the
1193 *Petunia* and the tomato/tobacco lineages where one *Petunia* loss event occurred (darker blue). If it was
1194 after the split, the presence of a *Petunia* gene would need to be explained by a gene gain through
1195 horizontal gene transfer or other means (cyan) - a far less likely scenario than a gene loss.

1196 **Figure 6 — figure supplement 5. Phylogenetic analysis of the BAHD acyltransferase.** The BAHD
1197 acyltransferase pseudogene (*Cc-BAHD-pseu*) in the corresponding syntenic region (Figure 6) of *Coffea*
1198 *canephora* is one of the closest *Coffea* sequences sister to the ASAT clade. It indicates that the BAHD

1199 acyltransferase gene was the first to harbor in this syntenic region before the divergence between
1200 Solanaceae and Rubiaceae. The translated amino acid sequence of *Cc-BAHD-pseu* was aligned with
1201 sequences used in Figure 6A of a previous study (Moghe et al., 2017) using MUSCLE. The phylogenetic
1202 trees were built using the maximum likelihood method with 1000 bootstrap replicates. The tree was
1203 generated using RAxML/8.0.6 with the following parameters: -f a -x 12345 -p 12345 -# 1000 -m
1204 PROTGAMMAAUTO --auto-prot=bic, and was shown with the midpoint rooting. Genes colored with green
1205 are from the focused syntenic regions. Genes marked with stars have been biochemically tested involved
1206 in acylsugar biosynthesis in previous studies.

1207 **Figure 6 — figure supplement 6. Additional evolutionary analysis of ACS genes in the syntenic**
1208 **regions to understand when the segmental duplication event happened.** (**A**) *Sl-AACS1* homologs
1209 obtained from *Salpiglossis sinuate* trichome transcriptome dataset (Moghe et al., 2017) were added for
1210 additional phylogenetic analysis. Genes colored with blue are from the syntenic regions of Solanaceae
1211 species Only the lineage derived by the number (4) evolutionary event as depicted in Figure 6 – figure
1212 supplement 5 was shown. (**B**) If the insertion happened before WGD, one gene loss on *Solanum*
1213 chromosome 12, as well as two independent gene losses on chromosomes 7 and 12 should have
1214 happened in *Petunia* and in *Salpiglossis sinuate* (Hypothesis 2). However, if the insertion happened after
1215 WGD, then only one gene loss event in *Petunia* and *Salpiglossis* was supposed to have happened
1216 (Hypothesis 1). The latter scenario is more likely due to the principle of parsimony.

1217 **Figure 6 — figure supplement 7. Ancestral trait state reconstruction analysis.** Four traits were
1218 inferred for their ancestral states using the maximum likelihood model Mk1 in Mesquite 3.6. They are the
1219 presence of medium chain acylsugars, presence of ACS in the synteny, presence of ECH in the synteny,
1220 and presence of both ACS and ECH in the synteny. The proportional likelihoods were shown in each
1221 diverging node in the ball charts.

1222 **Figure 6 — figure supplement 8. Phylogenetic distribution of acylsugar acyl chains with different**
1223 **lengths across the Solanaceae family.** (**A**) The collated results of acylsugar acyl chain distribution
1224 across different Solanaceae species. Red rectangles indicate detectable acyl chains in the acylsugars
1225 produced in the tested species and white rectangles indicate no detectable signals. The data source where
1226 the results are derived is listed on the right. (**B**) Results of acylsugar acyl chain characterization of selected

1227 Solanaceae species. Mole percentage (Mol %) of acylsugar acyl chains with different lengths were
1228 obtained from GC/MS analysis of fatty acid ethyl esters.

1229 **Supplementary file 1. Co-expression analysis of tomato genes from ACS, ECH, and BAHD**
1230 **acyltransferase families used for phylogenetic analysis in this study.** The values of Pearson's
1231 correlation coefficient of the expression profiles between any of the two genes were shown in the table.
1232 The coefficient values were generated using the FPKM values of these genes in the 372 RNA-seq
1233 samples as shown in Figure 2-supplemental figure1. The orange box highlights a group of co-expressed
1234 genes involved in acylsugar biosynthesis, such as *Sl-ASATs*, *Sl-AACS1*, and *Sl-AECH1*. The purple box
1235 points out another group of co-expressed genes that are root hair specific.

1236 **Supplementary file 2**. **Gene expression levels of all analyzed transcripts in *Solanum pennellii***
1237 **LA0716.** logFC: log2 fold change in stem trichomes versus shaved stems. logCPM: log (counts per million)
1238 in trichomes versus shaved stems. The F and Q-value test the significance of differential expression via a
1239 quasi- general linear model. The values noted in the sample columns represent the FPKM (Fragments Per
1240 Kilobase of transcript per Million mapped reads) analyzed via Cufflinks.

1241 **Supplementary file 3.** Synthesized gene fragments and primers used in this study.

1242 **Supplementary file 4.** The date used to generate the synteny figure shown in Figure 6 — figure
1243 supplement 1.

1244 **Supplementary file 5.** The sequence alignment documents used to generate the phylogenetic trees for
1245 Figure 6 — figure supplements 2, 3, 4, 5, 6.

1246 **Figure 2 — source data 1.** Data used to make Figure 2B. Peak area percentage of seven major trichome
1247 acylsugars in M82 and IL7-4.

1248 **Figure 3 — source data 1.** Data used to make Figure 3B. Quantification of seven major trichome
1249 acylsugars in the CRISPR mutants *sl-aacs1* and *sl-aech1*, as well as the parent M82.

1250 **Figure 4 — source data 1.** Data used to make Figure 4B. Aliphatic fatty acids of different chain lengths
1251 were used as the substrates to test Sl-AACS1 acyl-CoA synthase activity.

1252 **Figure 4 — source data 2.** Data used to make Figure 4D and Figure 4-figure supplement 1C. *N.*
1253 *benthamiana* leaf membrane lipid acyl chain composition.

1254 **Figure 5 — source data 1.** Data used to make Figure 5C and 5D. Silencing *Sq-AACS1* or *Sq-AECH1* in *S.*
1255 *quitoense* using VIGS leads to reduction of total acylsugars.
1256

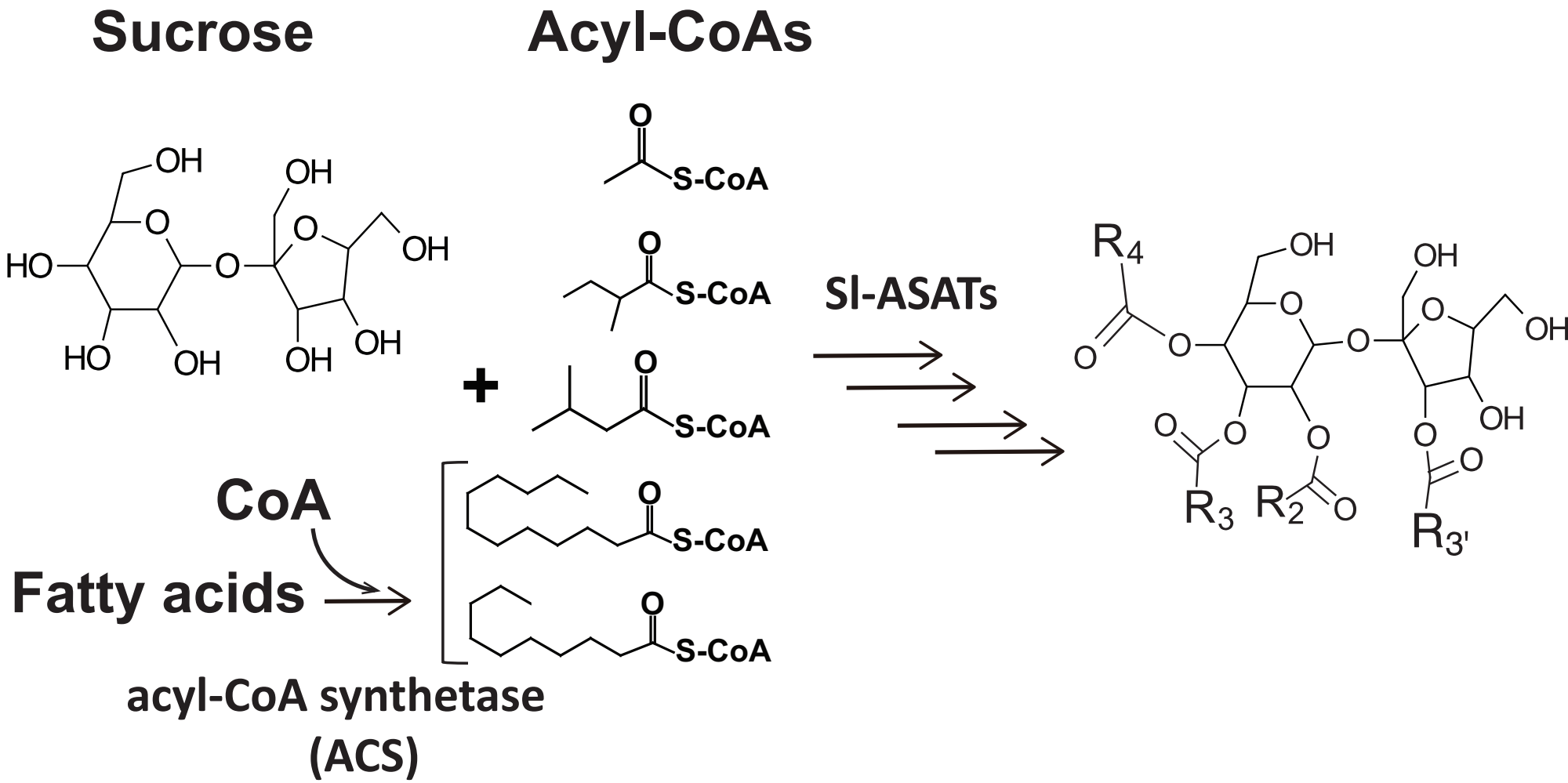

Sucrose
Acyl-CoAs
OH
OH
O
O
OH
HO
O
HO
OH
OH
OH
O
S-CoA
O
S-CoA
+
O
S-CoA
Sl-ASATs
R4
OH
OH
O
O
O
O
OH
O
O
O
O
O
O
R3
R2
R3'
CoA
Fatty acids
O
S-CoA
O
S-CoA
acyl-CoA synthetase
(ACS)

Figure 1

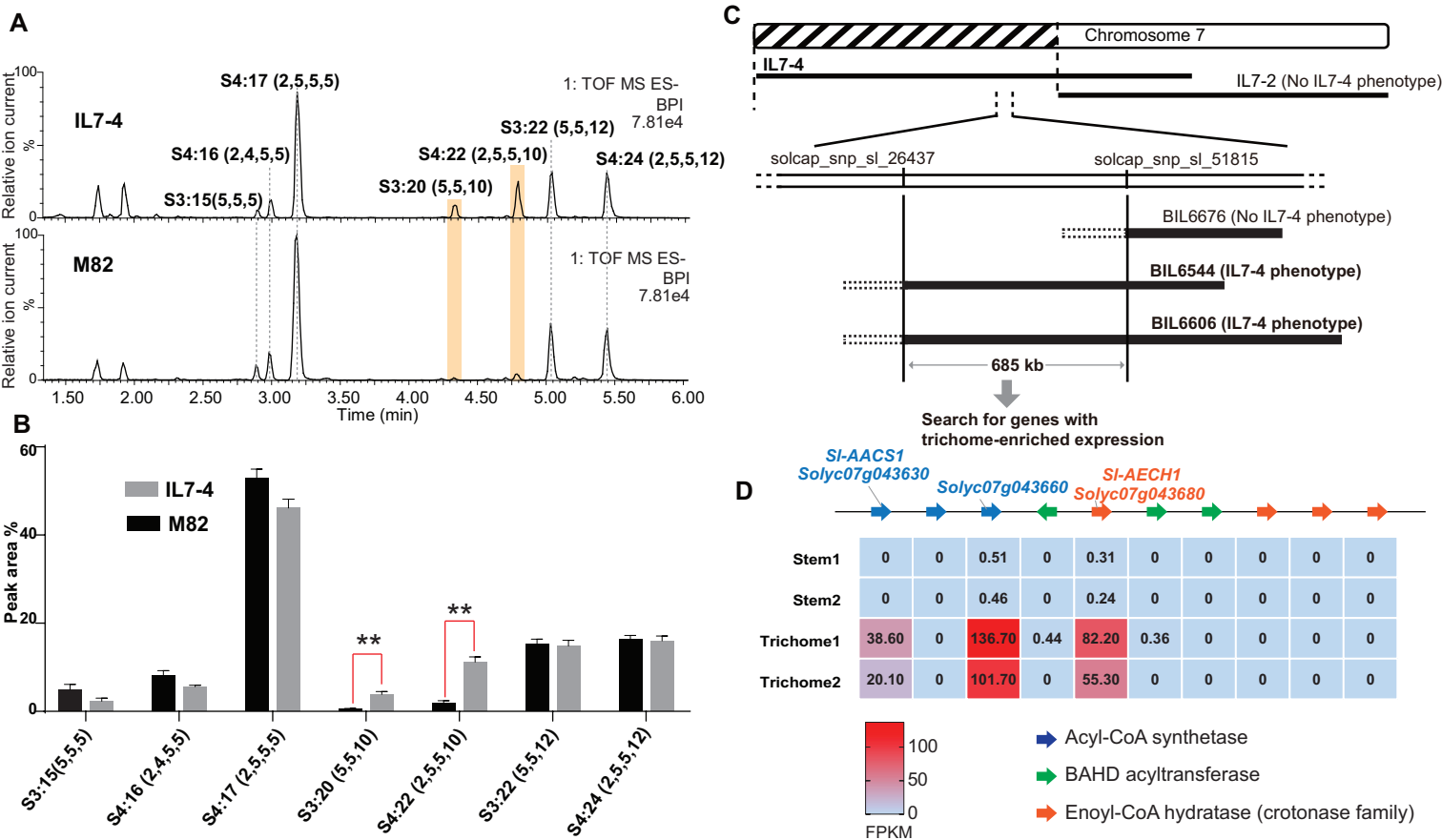

A
IL7-4
M82
Relative ion current
%
S4:17 (2,5,5,5)
S4:16 (2,4,5,5)
S3:15(5,5,5)
S3:20 (5,5,10)
S4:22 (2,5,5,10)
S3:22 (5,5,12)
S4:24 (2,5,5,12)
1: TOF MS ES-
BPI
7.81e4
Time (min)
B
Peak area %
IL7-4
M82
**
S3:15(5,5,5)
S4:16 (2,4,5,5)
S4:17 (2,5,5,5)
S3:20 (5,5,10)
S4:22 (2,5,5,10)
S3:22 (5,5,12)
S4:24 (2,5,5,12)
C
Chromosome 7
IL7-4
IL7-2 (No IL7-4 phenotype)
solcap_snp_sl_26437
solcap_snp_sl_51815
BIL6676 (No IL7-4 phenotype)
BIL6544 (IL7-4 phenotype)
BIL6606 (IL7-4 phenotype)
685 kb
Search for genes with
trichome-enriched expression
D
Sl-AACS1
Solyc07g043630
Solyc07g043660
Sl-AECH1
Solyc07g043680
Stem1 0 0 0.51 0 0.31 0 0 0 0 0
Stem2 0 0 0.46 0 0.24 0 0 0 0 0
Trichome1 38.60 0 136.70 0.44 82.20 0.36 0 0 0 0
Trichome2 20.10 0 101.70 0 55.30 0 0 0 0 0
FPKM
Acyl-CoA synthetase
BAHD acyltransferase
Enoyl-CoA hydratase (crotonase family)

Figure 2

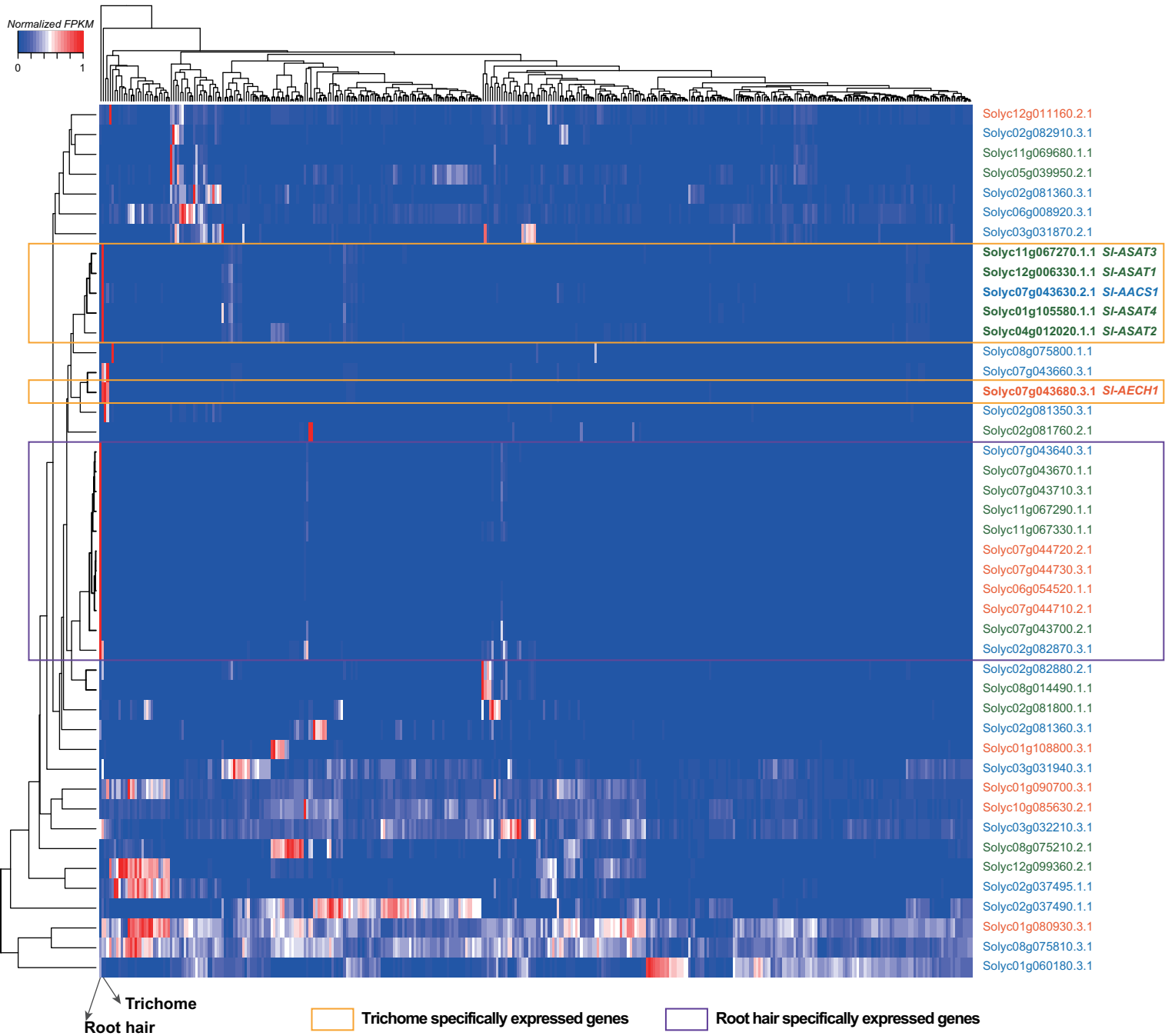


Figure 2 — figure supplement 1

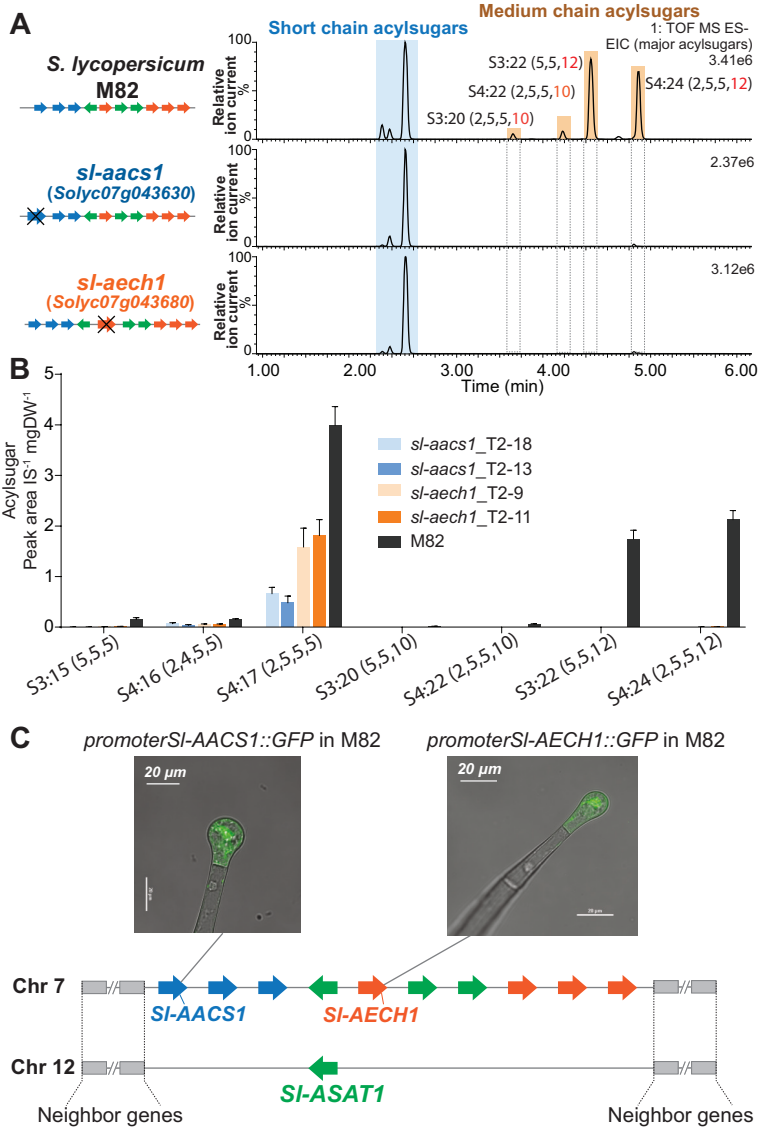

A
Short chain acylsugars
Medium chain acylsugars
1: TOF MS ES-
EIC (major acylsugars)
S. lycopersicum
M82
sl-aacs1
(Solyc07g043630)
sl-aech1
(Solyc07g043680)
Relative ion current %
3.41e6
2.37e6
3.12e6
S3:20 (2,5,5,10)
S4:22 (2,5,5,10)
S3:22 (5,5,12)
S4:24 (2,5,5,12)
Time (min)
B
Acylsugar Peak area IS-1 mgDW-1
sl-aacs1_T2-18
sl-aacs1_T2-13
sl-aech1_T2-9
sl-aech1_T2-11
M82
S3:15 (5,5,5)
S4:16 (2,4,5,5)
S4:17 (2,5,5,5)
S3:20 (5,5,10)
S4:22 (2,5,5,10)
S3:22 (5,5,12)
S4:24 (2,5,5,12)
C
promoterSl-AACS1::GFP in M82
promoterSl-AECH1::GFP in M82
20 µm
Chr 7
Sl-AACS1
Sl-AECH1
Chr 12
Sl-ASAT1
Neighbor genes
Neighbor genes

Figure 3

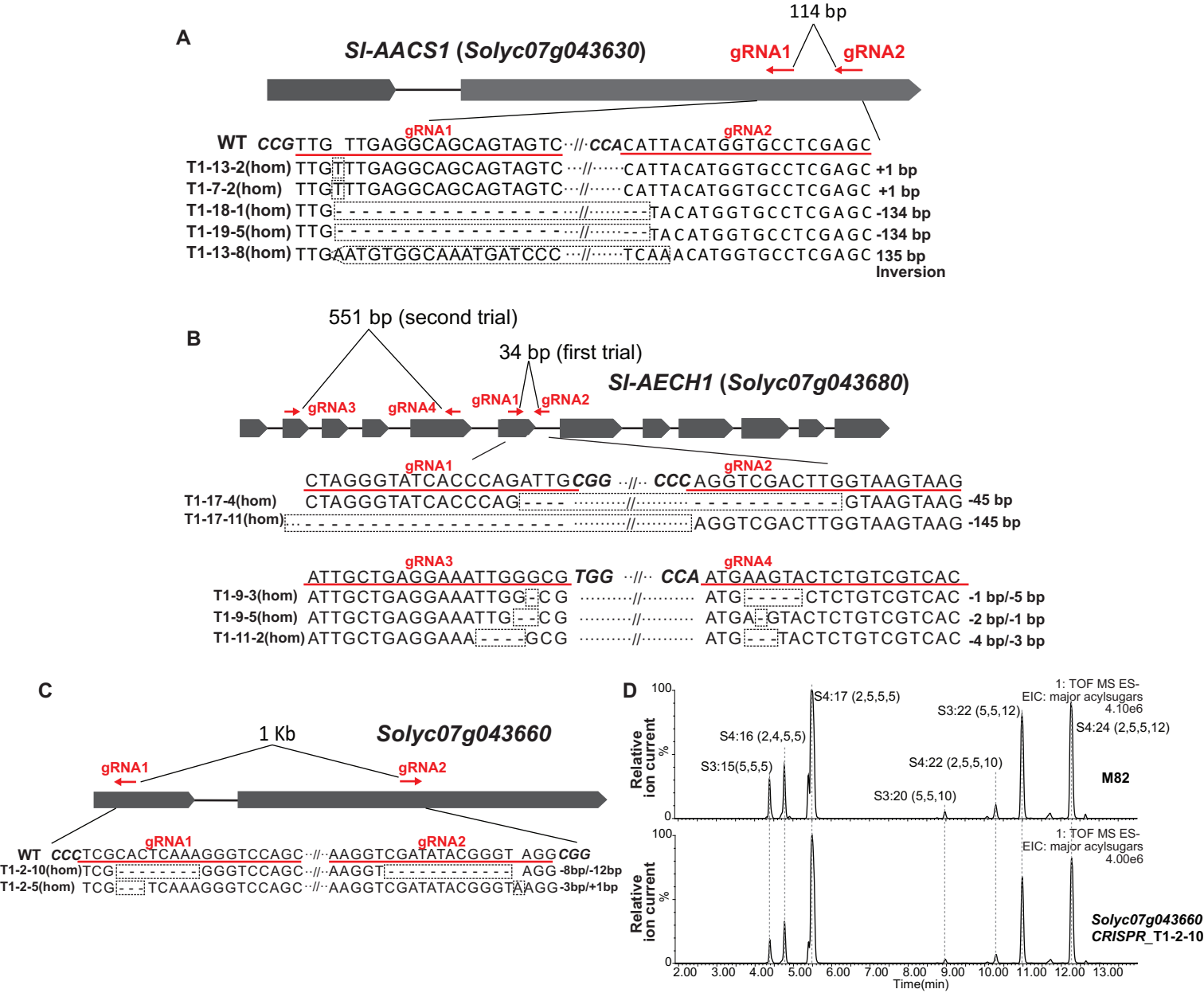


Figure 3 — figure supplement 1

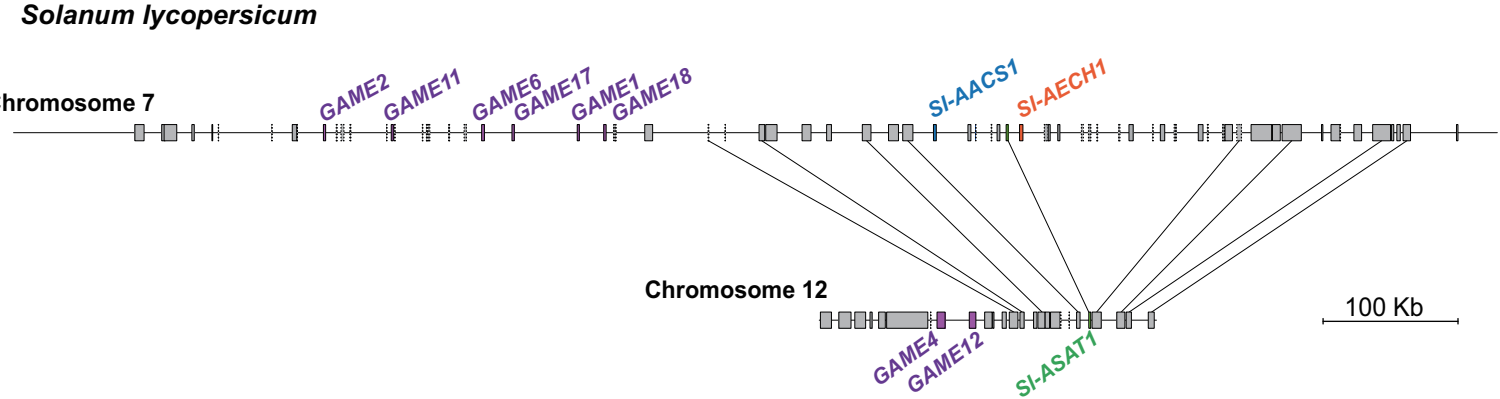
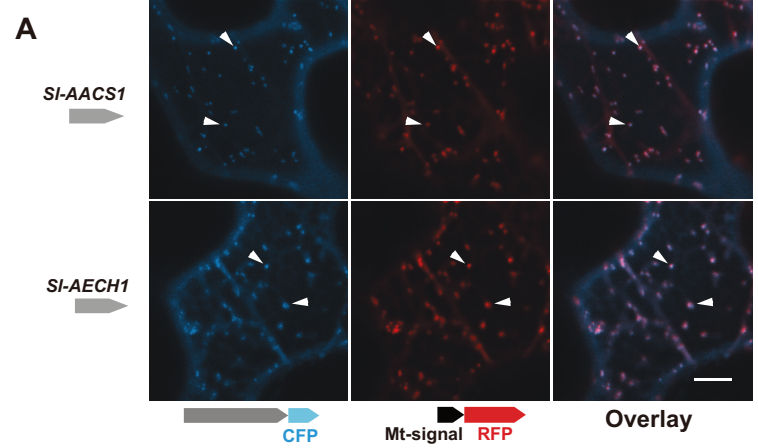
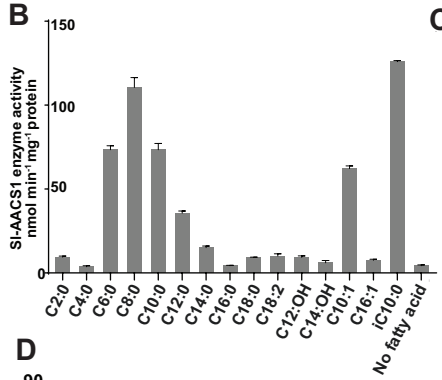


Figure 3 — figure supplement 2





C

	K_m (μM)	k_{cat} (min^{-1})	k_{cat}/K_m ($\mu M^{-1}/min^{-1}$)
C6:0	214.8 ± 40	13.8 ± 1.2	0.064
C8:0	145.9 ± 10.4	13.2 ± 0.4	0.09
C10:0	71.8 ± 9.6	5.6 ± 0.2	0.078
C12:0	68.1 ± 15.8	4.2 ± 0.3	0.062
iC10:0	82.1 ± 10.3	13.6 ± 0.6	0.166
C10:1	36.4 ± 4.5	5.9 ± 0.2	0.162

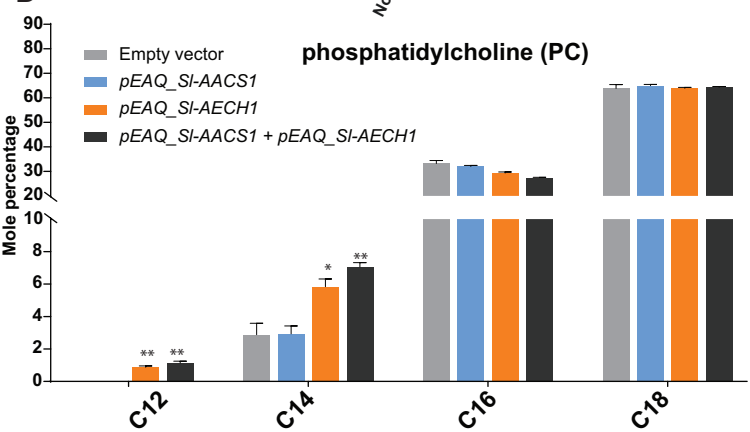


Figure 4

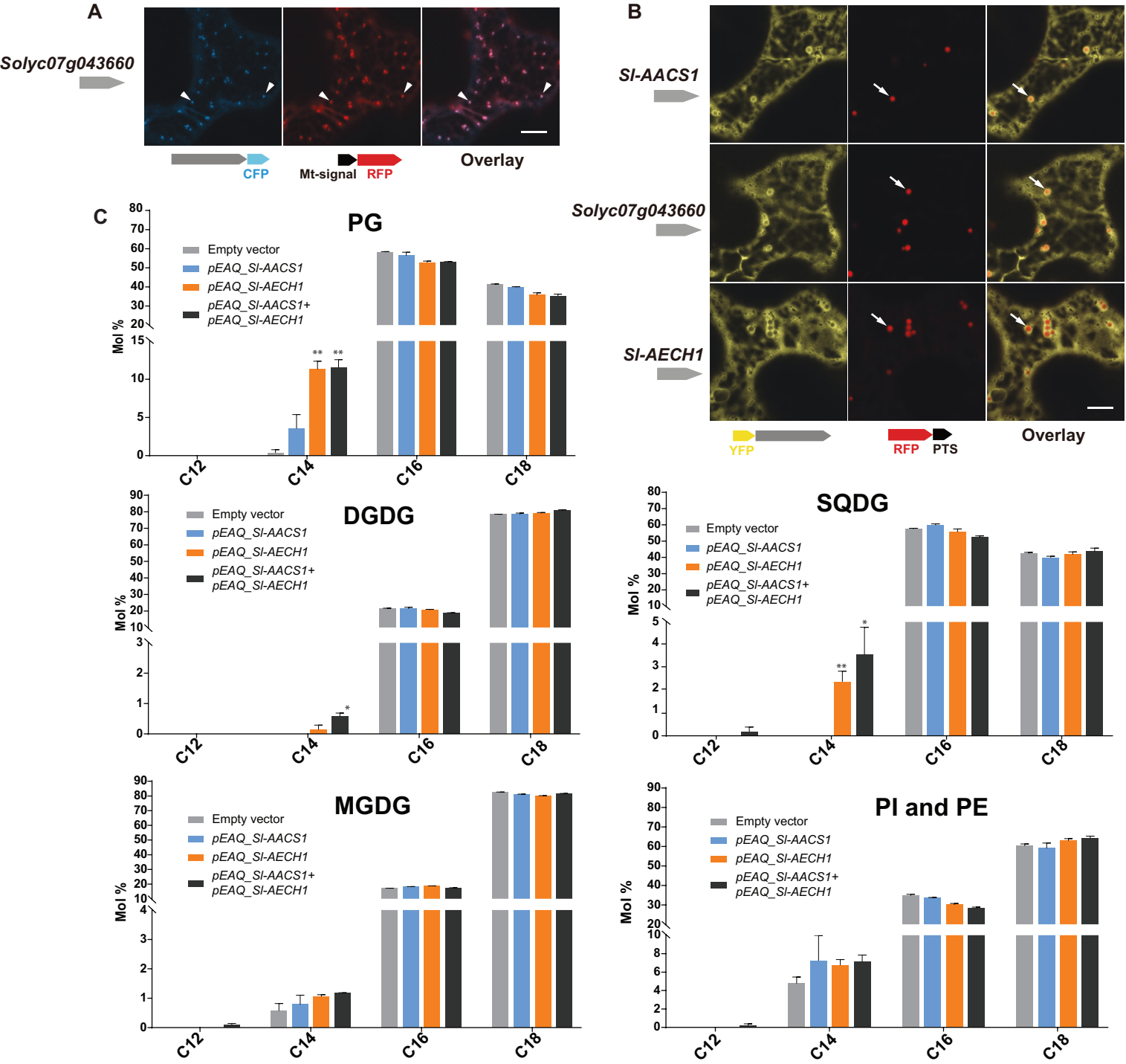


Figure 4 — figure supplement 1

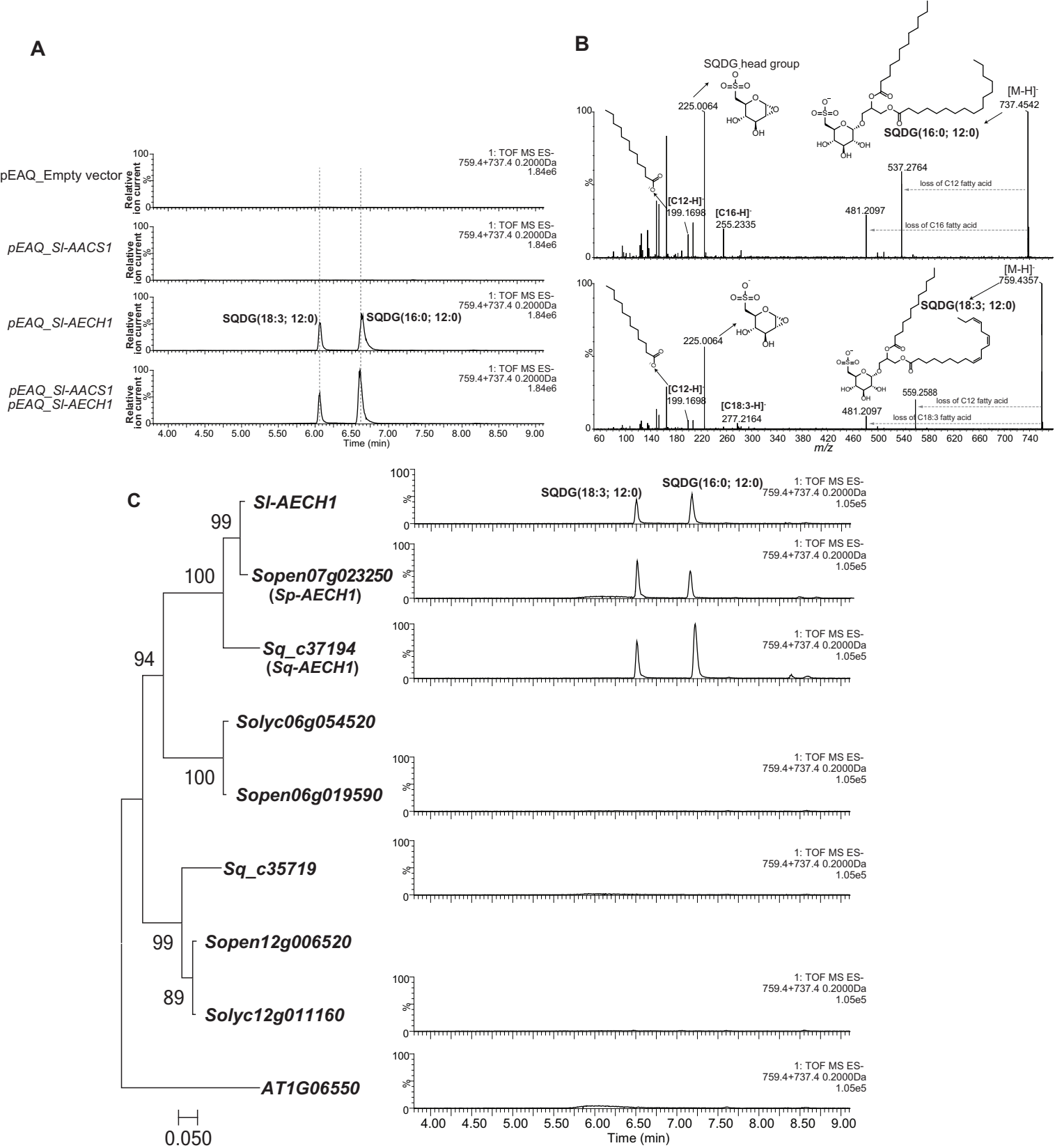


Figure 4 — figure supplement 2

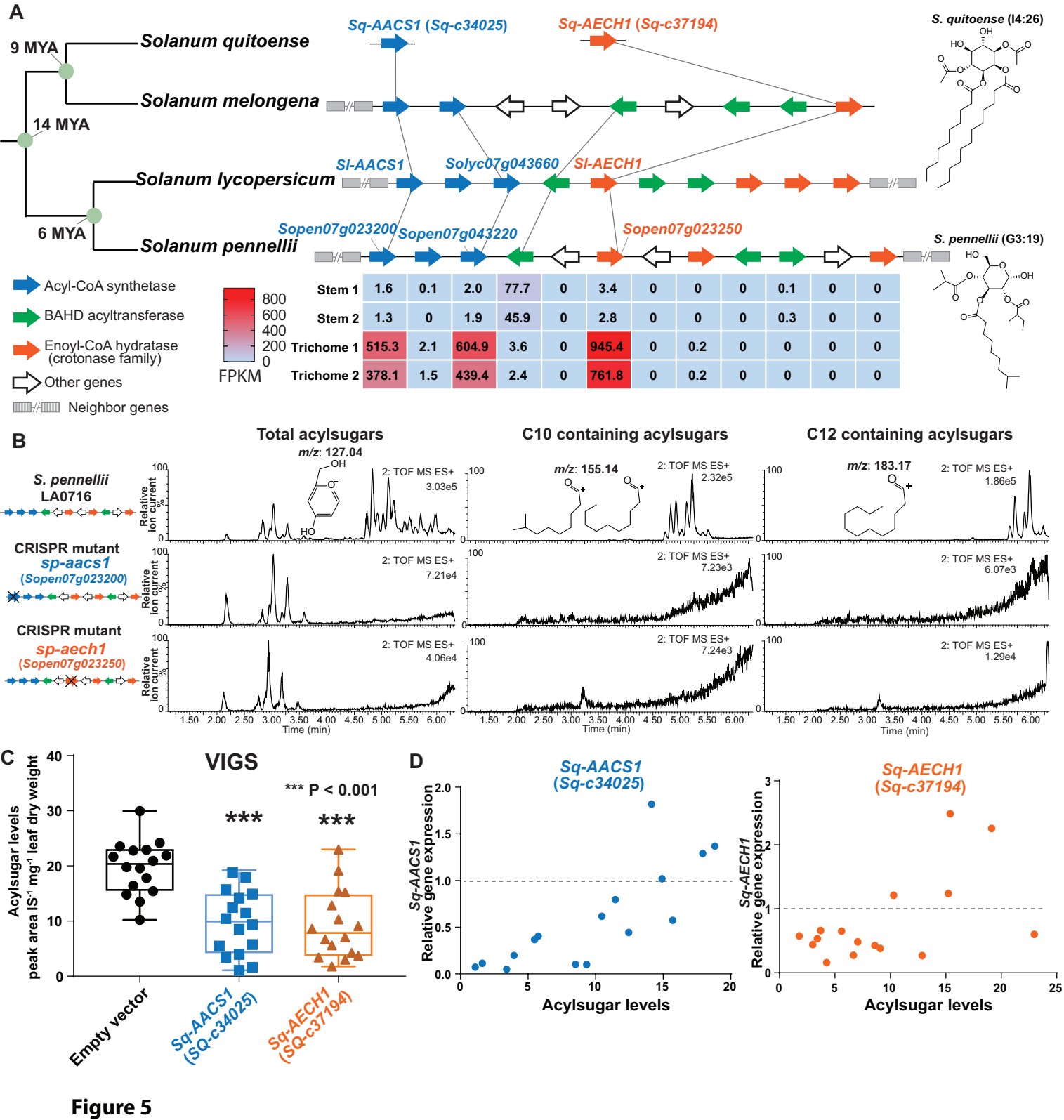

A
9 MYA
14 MYA
6 MYA
Solanum quitoense
Solanum melongena
Solanum lycopersicum
Solanum pennellii
Sq-AACS1 (Sq-c34025)
Sq-AECH1 (Sq-c37194)
Sl-AACS1
Solyc07g043660
Sl-AECH1
Sopen07g023200
Sopen07g043220
Sopen07g023250
S. quitoense (I4:26)
S. pennellii (G3:19)
Acyl-CoA synthetase
BAHD acyltransferase
Enoyl-CoA hydratase (crotonase family)
Other genes
Neighbor genes
800
600
400
200
0
FPKM
Stem 1 1.6 0.1 2.0 77.7 0 3.4 0 0 0 0.1 0 0
Stem 2 1.3 0 1.9 45.9 0 2.8 0 0 0 0.3 0 0
Trichome 1 515.3 2.1 604.9 3.6 0 945.4 0 0.2 0 0 0 0
Trichome 2 378.1 1.5 439.4 2.4 0 761.8 0 0.2 0 0 0 0
B
Total acylsugars
C10 containing acylsugars
C12 containing acylsugars
m/z: 127.04
m/z: 155.14
m/z: 183.17
S. pennellii LA0716
CRISPR mutant sp-aacs1 (Sopen07g023200)
CRISPR mutant sp-aech1 (Sopen07g023250)
2: TOF MS ES+
3.03e5
7.21e4
4.06e4
2.32e5
7.23e3
7.24e3
1.86e5
6.07e3
1.29e4
Relative ion current %
Time (min)
C
VIGS
*** P < 0.001
***
***
Acylsugar levels peak area IS-1 mg-1 leaf dry weight
Empty vector
Sq-AACS1 (SQ-c34025)
Sq-AECH1 (SQ-c37194)
D
Sq-AACS1 (Sq-c34025)
Sq-AECH1 (Sq-c37194)
Sq-AACS1 Relative gene expression
Sq-AECH1 Relative gene expression
Acylsugar levels

Figure 5

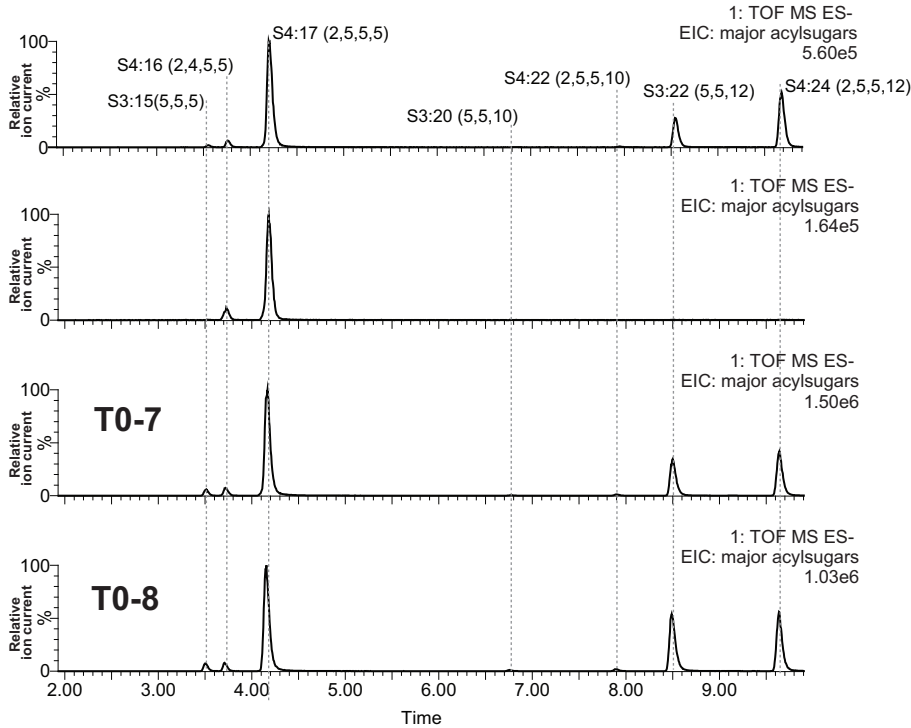
A

S. lycopersicum M82

*sl-aacs1*_T2-13-2

promoterSopen07g023200::Sopen07g023200 in *sl-aacs1*

promoterSopen07g023200::Sopen07g023200 in *sl-aacs1*



B

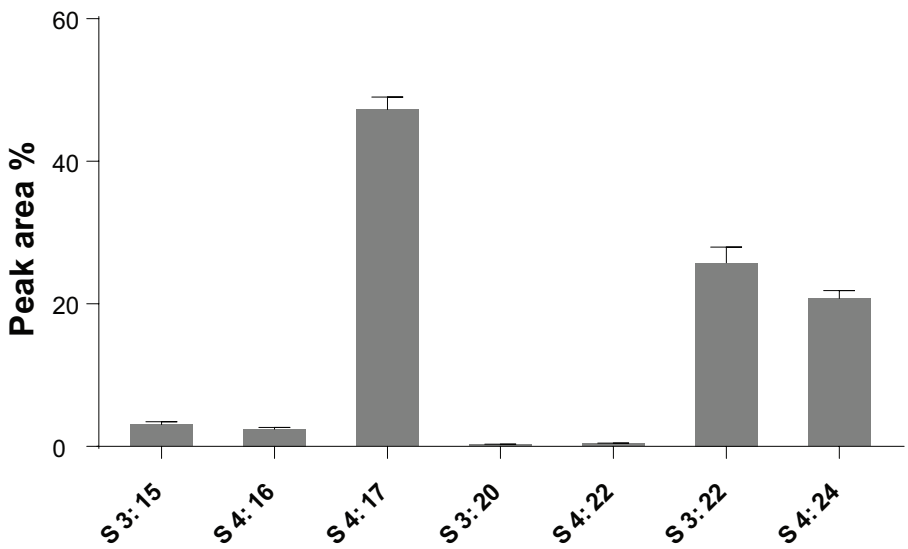


Figure 5 — figure supplement 1

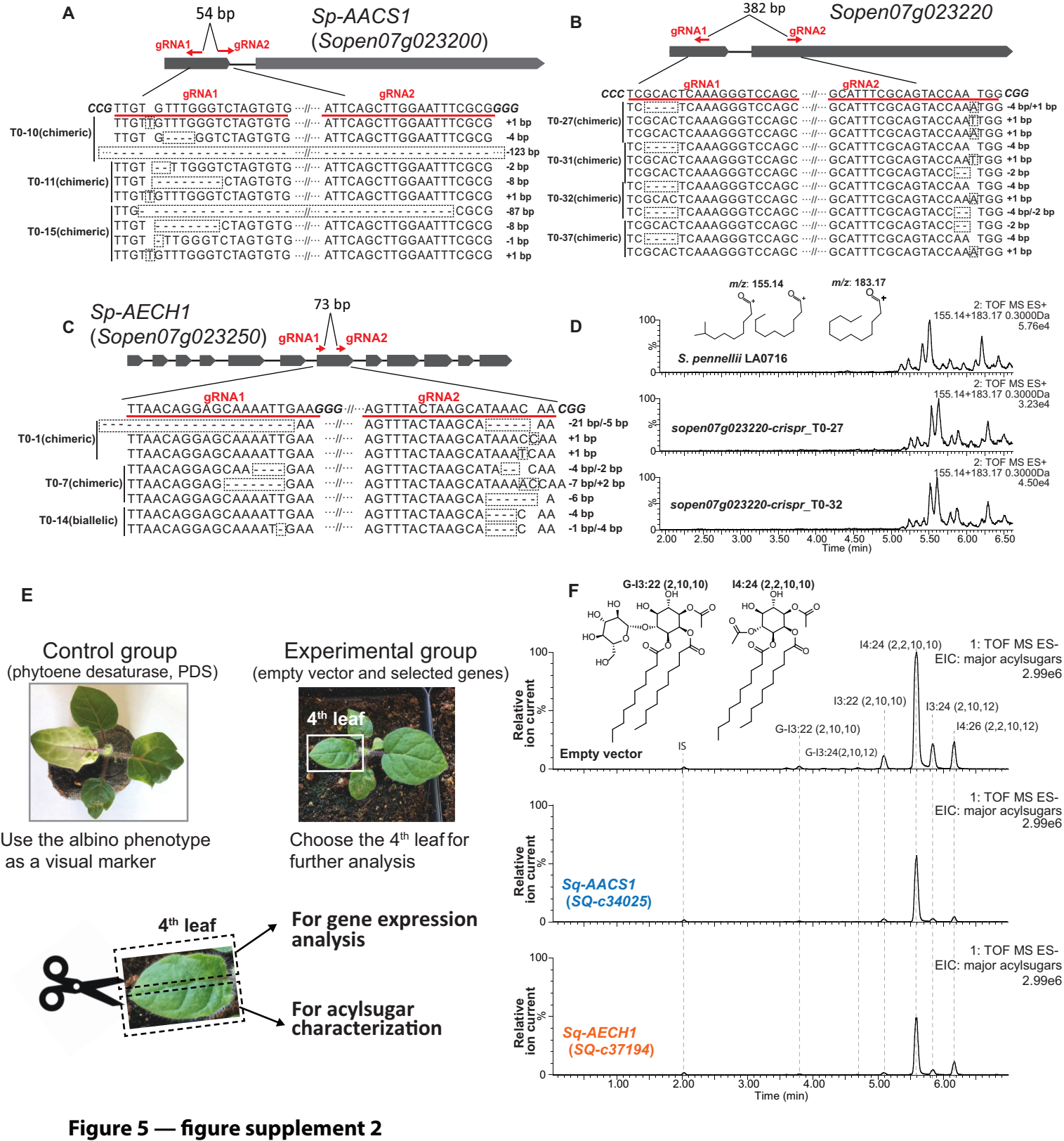


Figure 5 — figure supplement 2

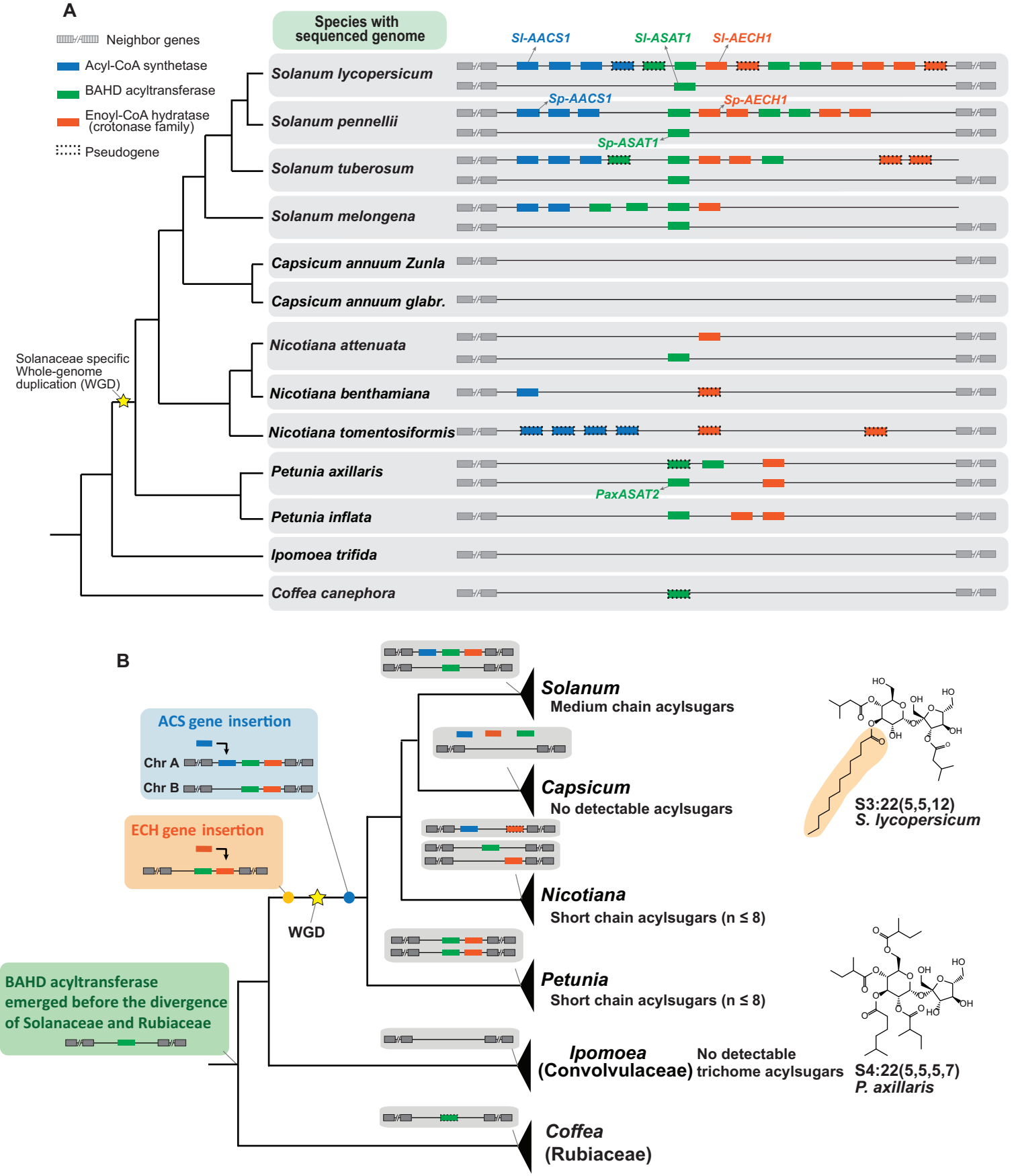

A
Neighbor genes
Acyl-CoA synthetase
BAHD acyltransferase
Enoyl-CoA hydratase (crotonase family)
Pseudogene
Species with sequenced genome
Sl-AACS1
Sl-ASAT1
Sl-AECH1
Solanum lycopersicum
Sp-AACS1
Sp-AECH1
Solanum pennellii
Sp-ASAT1
Solanum tuberosum
Solanum melongena
Capsicum annuum Zunla
Capsicum annuum glabr.
Nicotiana attenuata
Nicotiana benthamiana
Nicotiana tomentosiformis
Petunia axillaris
PaxASAT2
Petunia inflata
Ipomoea trifida
Coffea canephora
Solanaceae specific Whole-genome duplication (WGD)
B
ACS gene insertion
Chr A
Chr B
ECH gene insertion
WGD
BAHD acyltransferase emerged before the divergence of Solanaceae and Rubiaceae
Solanum
Medium chain acylsugars
Capsicum
No detectable acylsugars
Nicotiana
Short chain acylsugars (n ≤ 8)
Petunia
Short chain acylsugars (n ≤ 8)
Ipomoea
(Convolvulaceae)
No detectable trichome acylsugars
Coffea
(Rubiaceae)
S3:22(5,5,12)
S. lycopersicum
S4:22(5,5,5,7)
P. axillaris

Figure 6

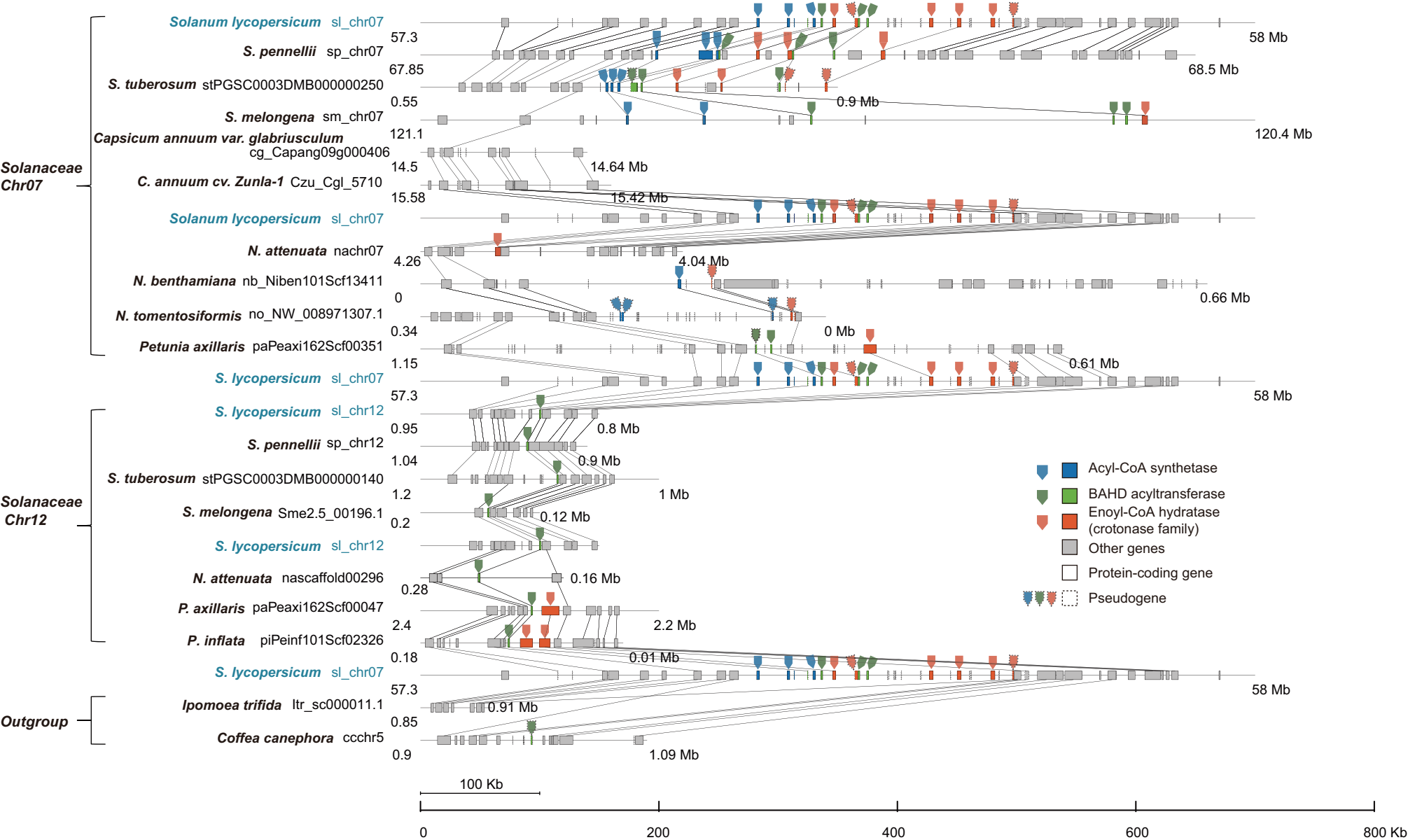


Figure 6 — figure supplement 1

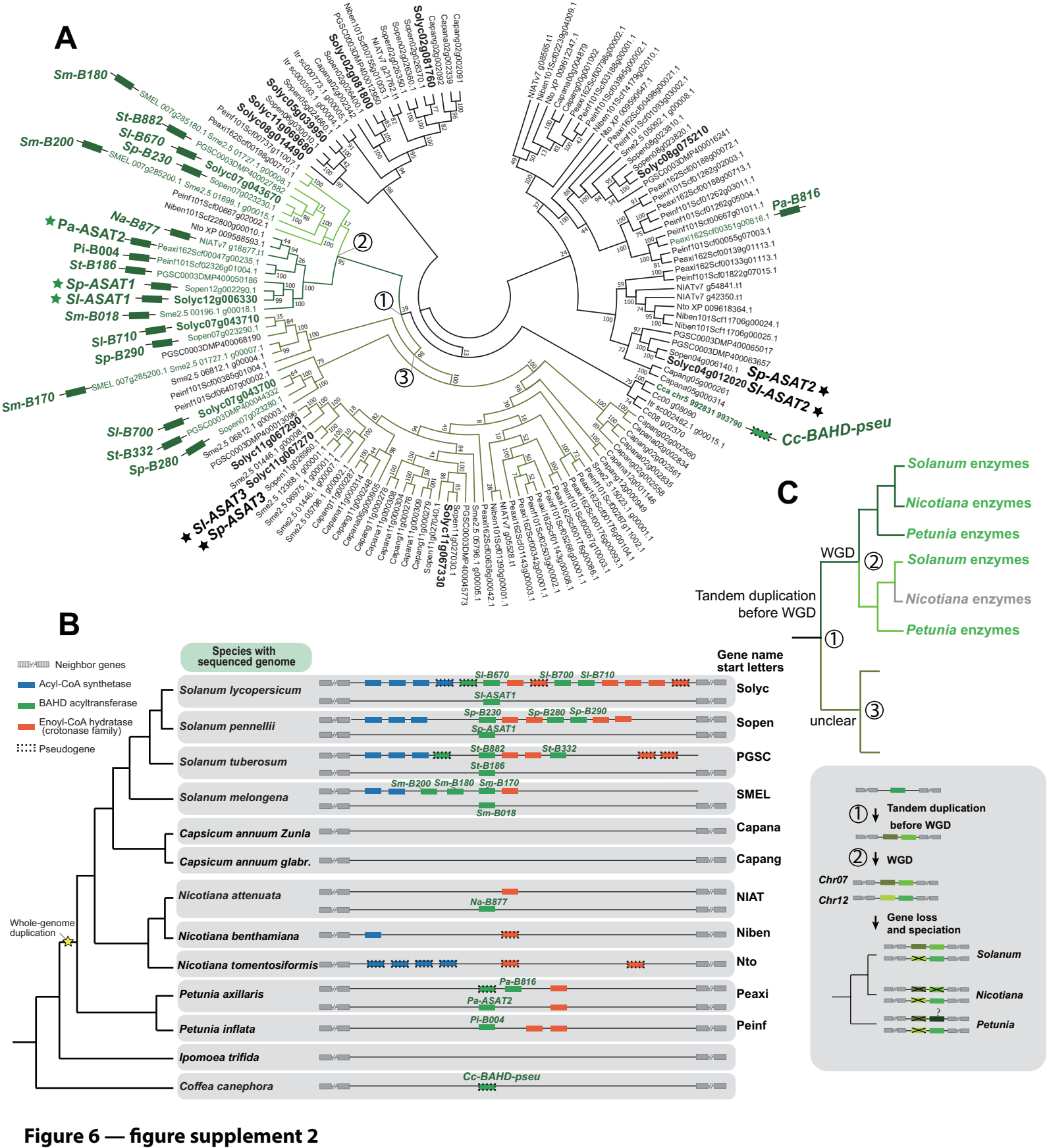


Figure 6 — figure supplement 2

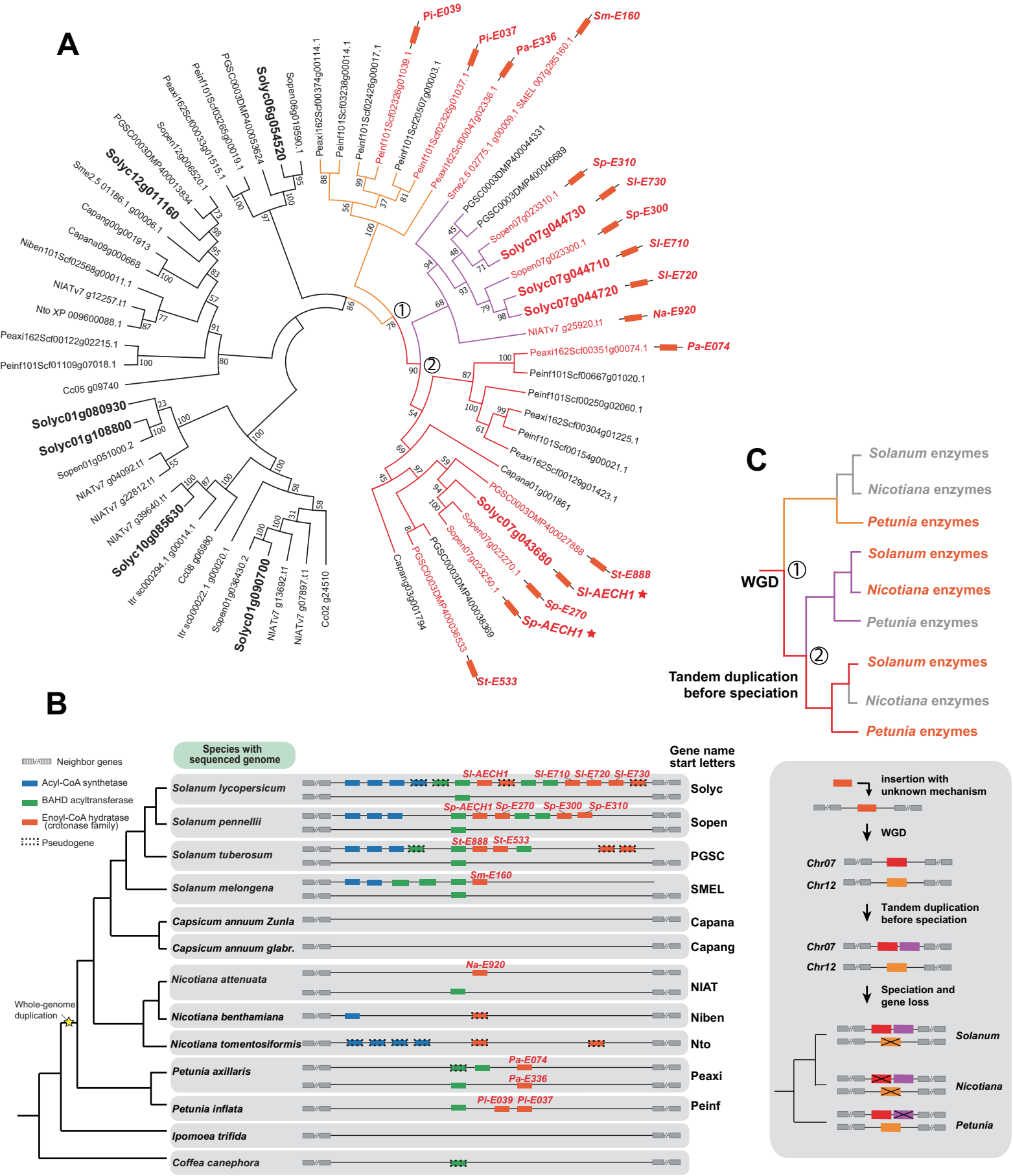


Figure 6 — figure supplement 3

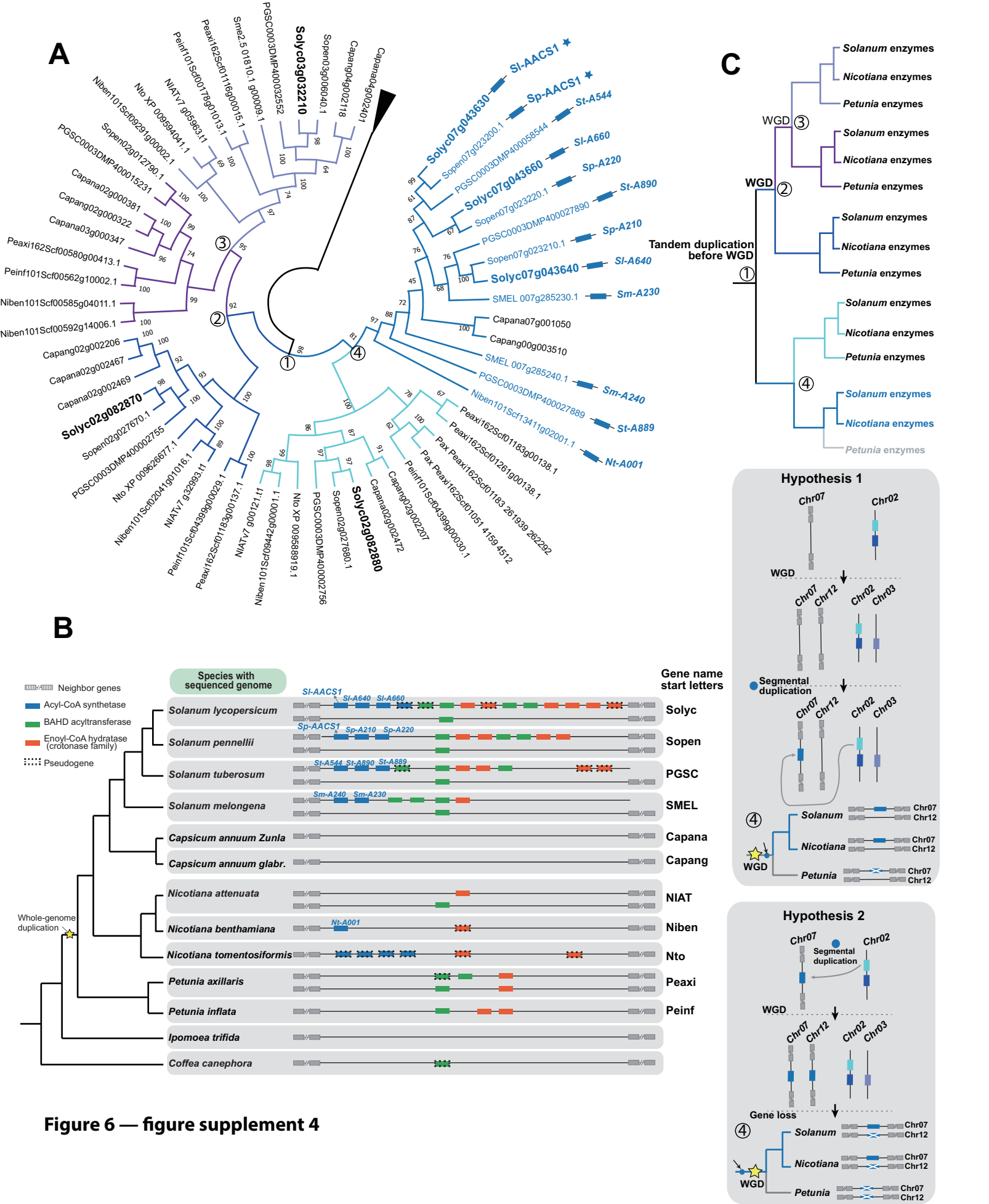


Figure 6 — figure supplement 4

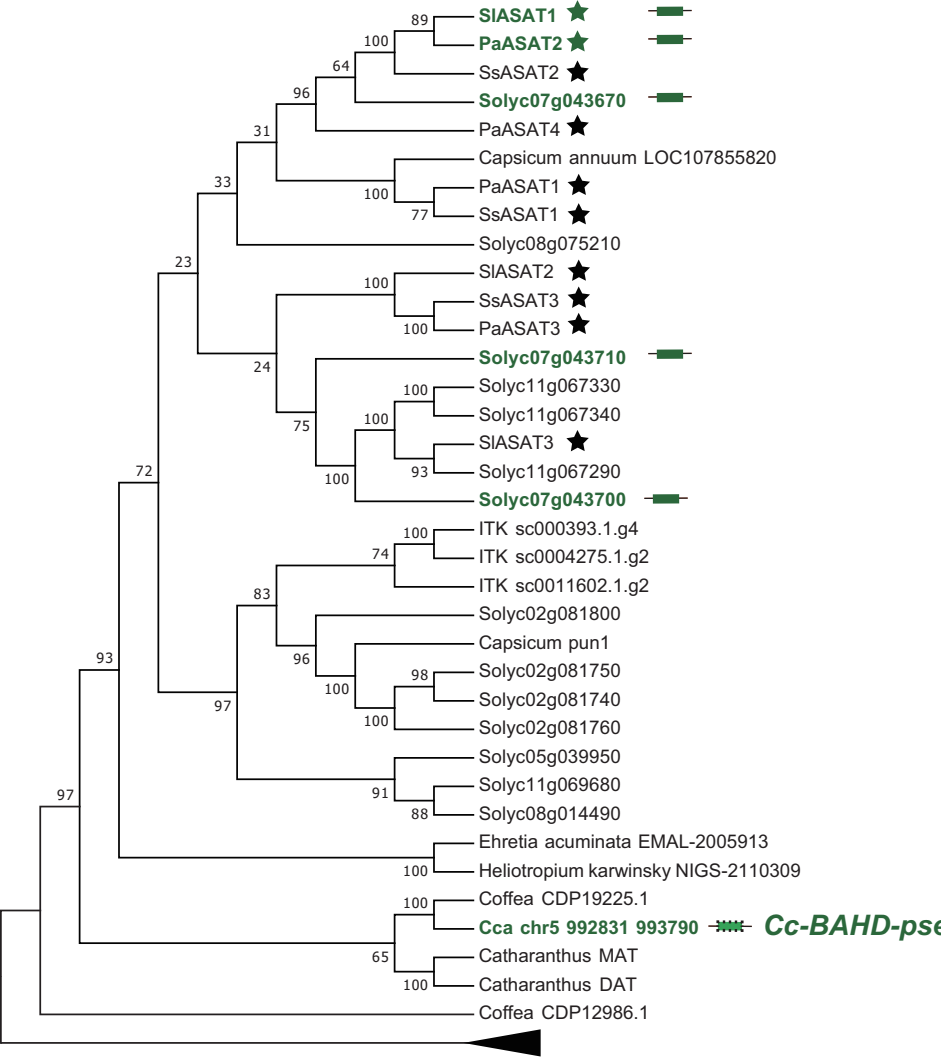


Figure 6 — figure supplement 5

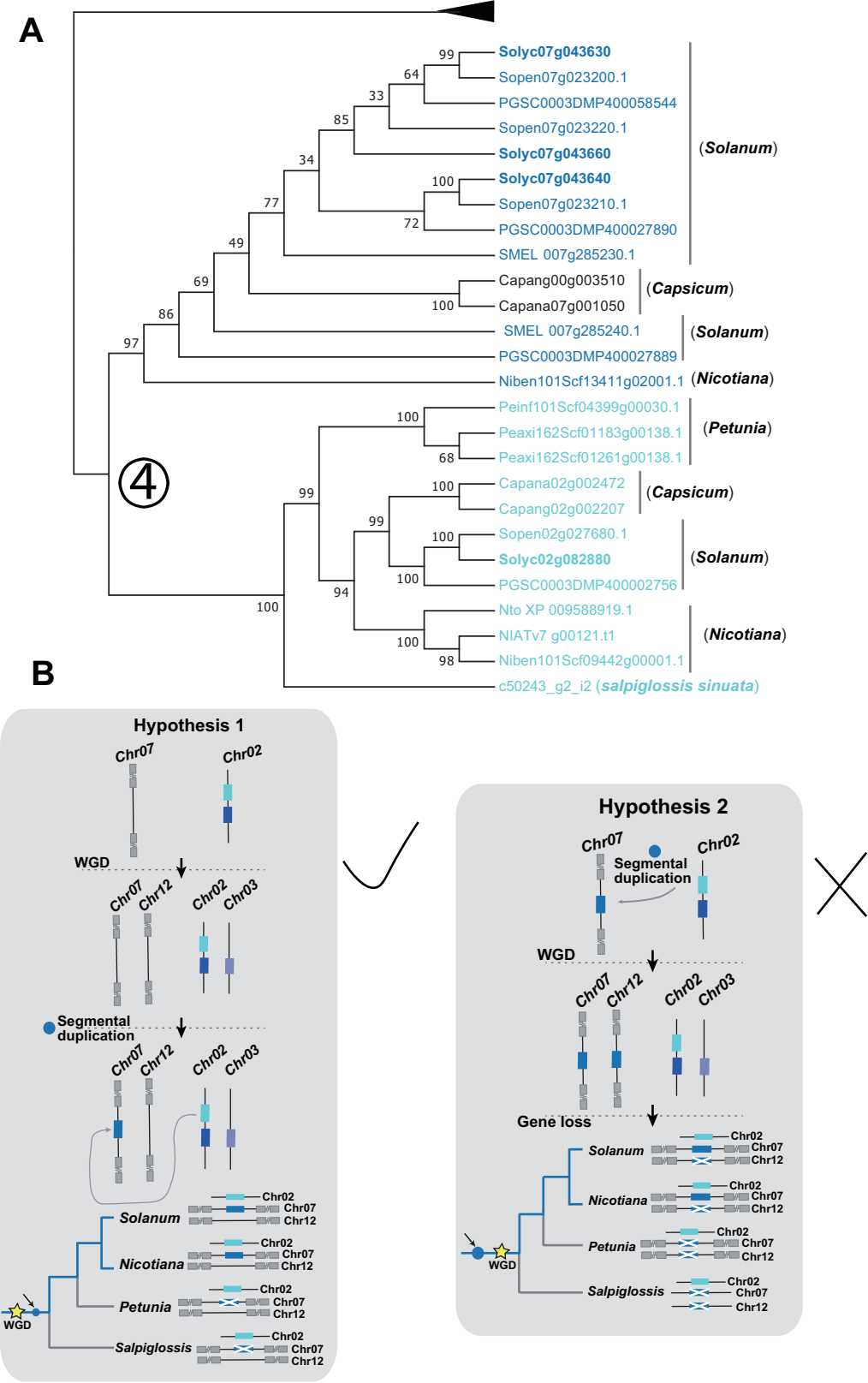


Figure 6 — figure supplement 6

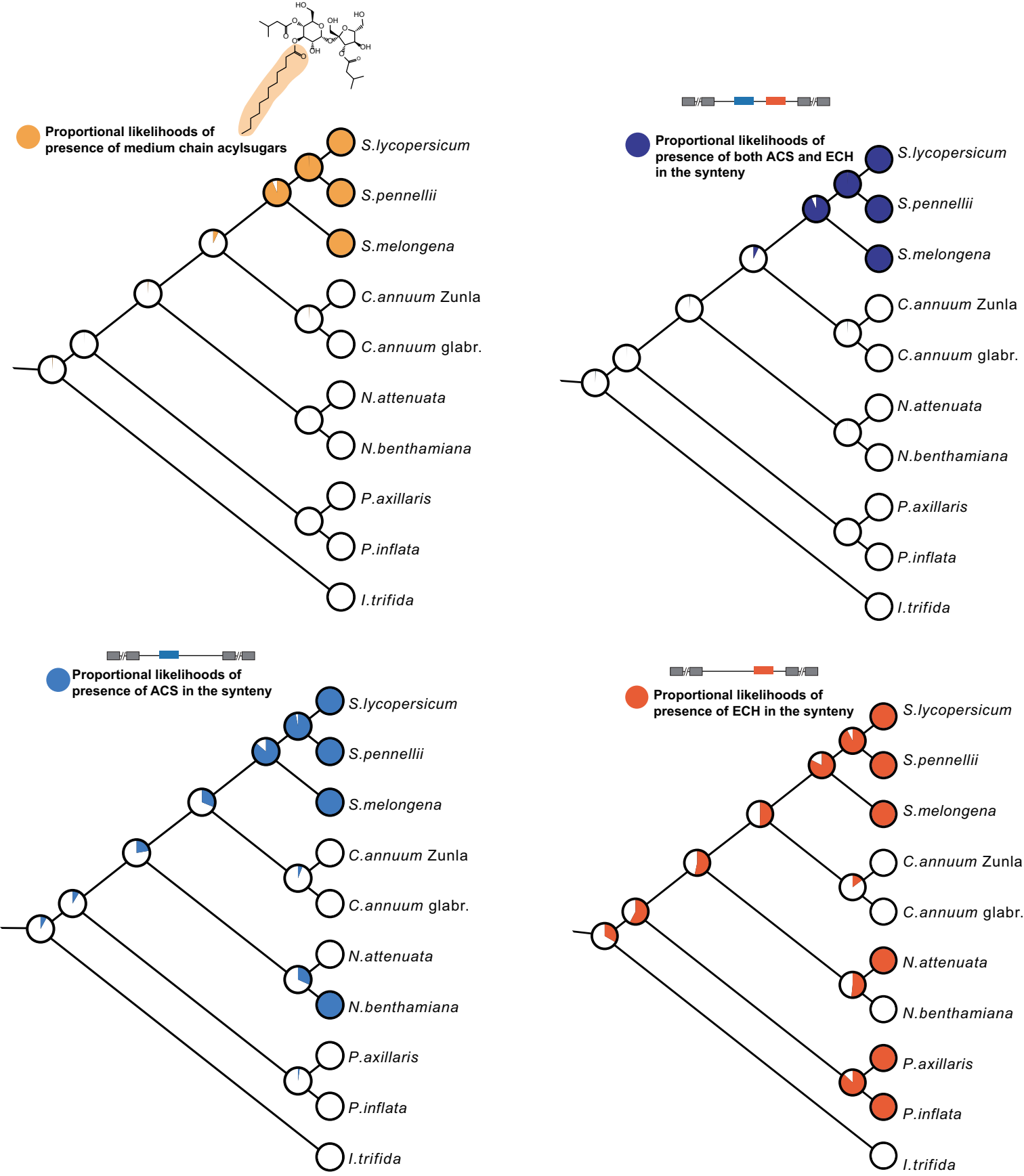
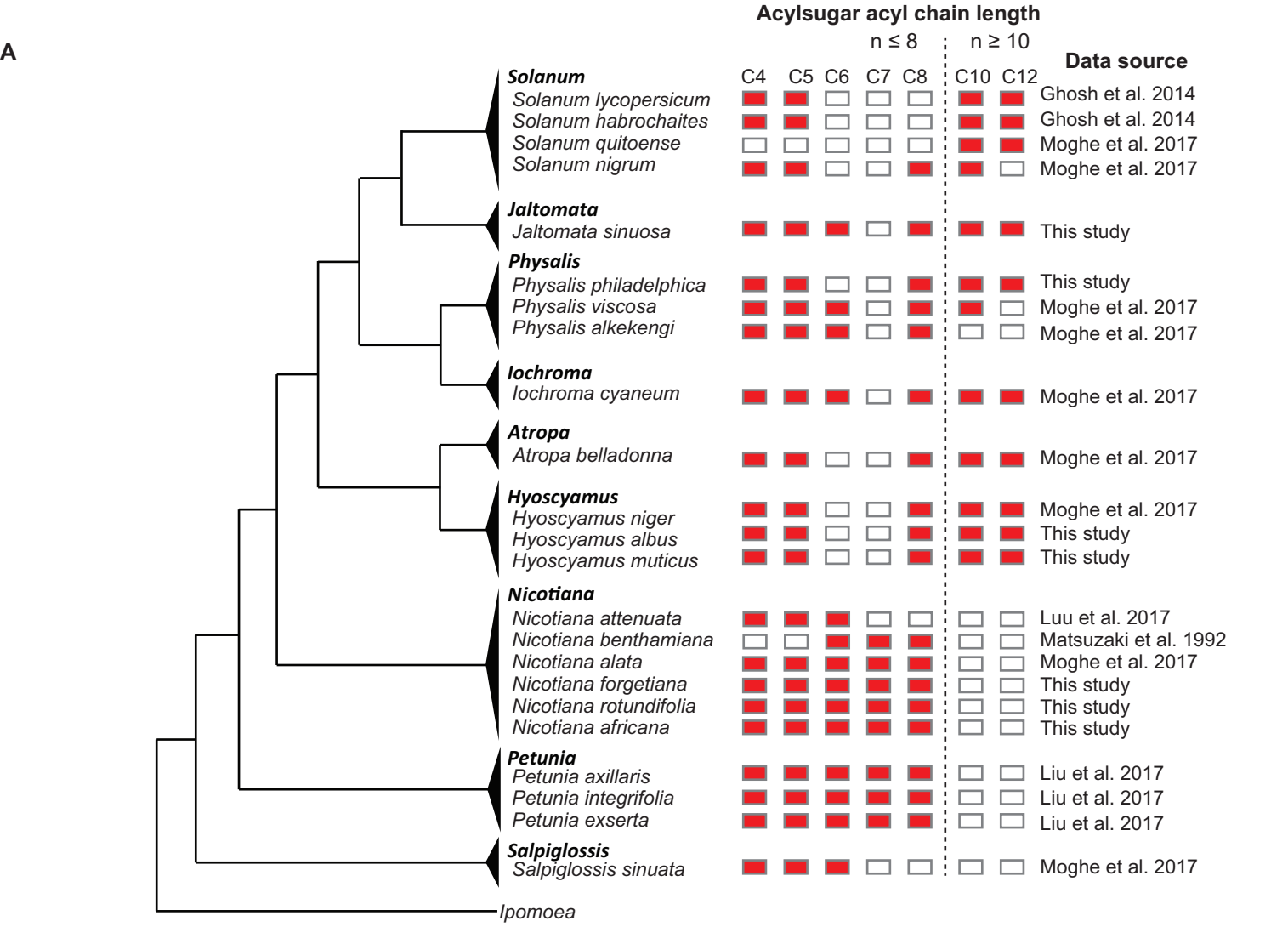


Figure 6 — figure supplement 7



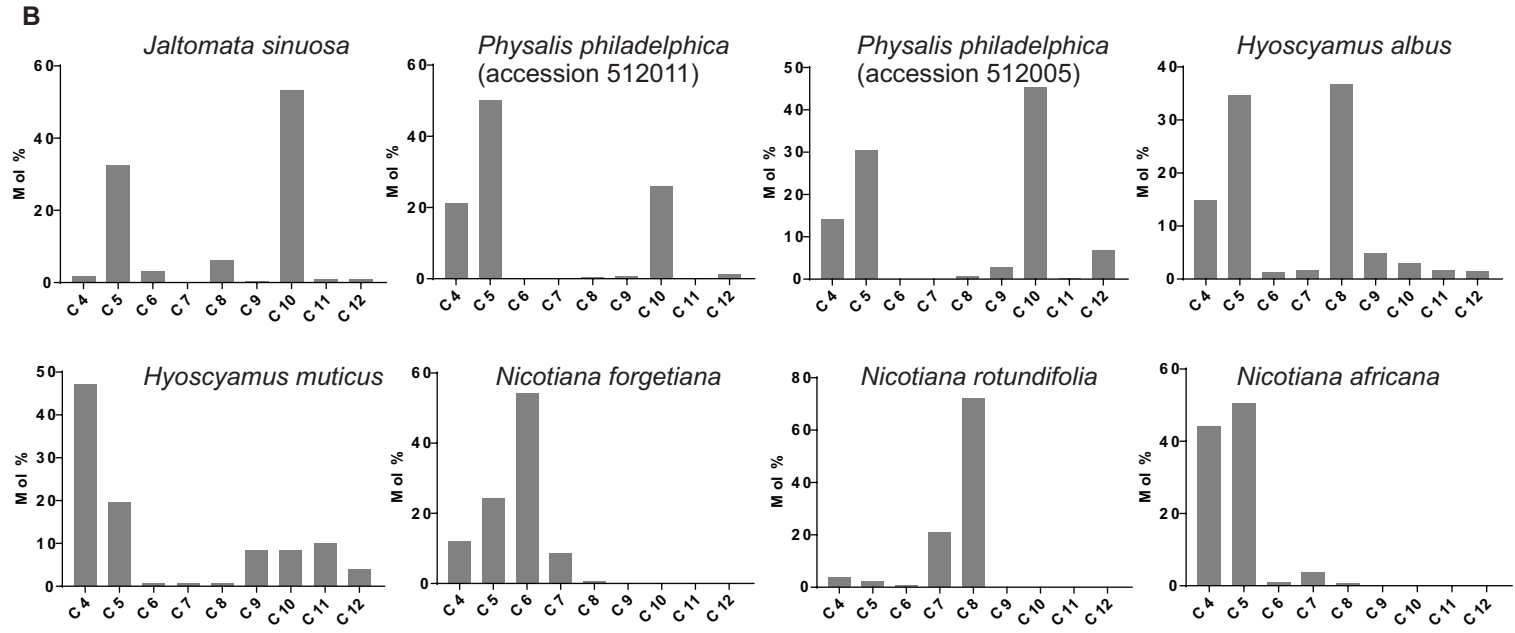


Figure 6 — figure supplement 8