Natural Product Reports

Fruity, Sticky, Stinky, Spicy, Bitter, Addictive, and Deadly: Evolutionary Signatures of Metabolic Complexity in the Solanaceae

Journal:	*Natural Product Reports*
Manuscript ID	NP-REV-01-2022-000003.R1
Article Type:	Review Article
Date Submitted by the Author:	01-Mar-2022
Complete List of Authors:	Fiesel, Paul; Michigan State University Parks, Hannah; Michigan State University Last, Robert; Michigan State University Barry, Cornelius; Michigan State University

Fruity, Sticky, Stinky, Spicy, Bitter, Addictive, and Deadly: Evolutionary Signatures of Metabolic Complexity in the Solanaceae

Paul D. Fiesel[a*], Hannah M. Parks[a*], Robert L. Last[a,b], Cornelius S. Barry[c†]

[a] Department of Biochemistry & Molecular Biology, Michigan State University, East Lansing, MI 48824

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

[c] Department of Horticulture, Michigan State University, East Lansing, MI 48824

* These authors contributed equally to this review

[†] For correspondence barrycs@msu.edu

ORCID numbers

0000-0002-2038-5417 – Paul D. Fiesel

0000-0002-9861-8049 – Hannah M. Parks

0000-0001-6974-9587—Robert L. Last

0000-0003-4685-0273 – Cornelius S. Barry

Abstract

Plants collectively synthesize a huge repertoire of metabolites. General metabolites, also referred to as primary metabolites, are conserved across the plant kingdom and are required for processes essential to growth and development. These include amino acids, sugars, lipids, and organic acids. In contrast, specialized metabolites, historically termed secondary metabolites, are structurally diverse, exhibit lineage-specific distribution and provide selective advantage to host species to facilitate reproduction and environmental adaptation. Due to their potent bioactivities, plant specialized metabolites attract considerable attention for use as flavorings, fragrances, pharmaceuticals, and bio-pesticides. The Solanaceae (Nightshade family) consists of approximately 2700 species and includes crops of significant economic, cultural, and scientific importance: these include potato, tomato, pepper, eggplant, tobacco, and petunia. The Solanaceae has emerged as a model family for studying the biochemical evolution of plant specialized metabolism and multiple examples exist of lineage-specific metabolites that influence the senses and physiology of commensal and harmful organisms, including humans. These include, alcohols, phenylpropanoids, and carotenoids that contribute to fruit aroma and color in tomato (***fruity***), glandular trichome-derived terpenoids and acylsugars that contribute to plant defense (***stinky*** & ***sticky***, respectively), capsaicinoids in chilli-peppers that influence seed dispersal (***spicy***), and steroidal glycoalkaloids (***bitter***) from *Solanum*, nicotine (***addictive***) from tobacco, as well as tropane alkaloids (***deadly***) from Deadly Nightshade that deter herbivory. Advances in genomics and metabolomics, coupled with the adoption of comparative phylogenetic approaches, resulted in deeper knowledge of the biosynthesis and evolution of these metabolites. This review highlights recent progress in this area and outlines opportunities for - and challenges of-developing a more comprehensive understanding of Solanaceae metabolism.

61

62 **1. The Solanaceae: a phylogenetic framework for exploring metabolism**

63 Metabolism is a window into micro- and macro-evolutionary processes. Plant metabolic diversity is vast
64 and collectively plants are hypothesized to synthesize ~10^6 metabolites[1]. Many of these metabolites,
65 including sugars, amino acids, fatty acids, and organic acids - referred to as general or primary metabolites
66 - are conserved across the plant kingdom, and essential for growth and development. However,
67 specialized metabolites (SM), also referred to in the literature as secondary metabolites, comprise the
68 majority of plant metabolic complexity. Specialized metabolites are chemically diverse, display
69 taxonomically restricted distribution, and are often synthesized in individual tissues or cell types. Plants
70 evolved the capacity to synthesize specific classes of specialized metabolites to facilitate ecological
71 adaptations. The advent of genomics, coupled with the ability to test the function of candidate genes in
72 host species or heterologous systems, advanced our understanding of the biosynthesis and evolution of
73 plant specialized metabolism[2-4].

74 Although plant specialized metabolites exhibit considerable chemical complexity, they are ultimately
75 derived from a pool of general metabolites formed through photosynthesis, glycolysis, the TCA cycle,
76 amino acid metabolism and the MEP-pathway[5]. General metabolites undergo transformations, including
77 ligation and cyclization to generate scaffold molecules that are modified by glycosylation, acylation,
78 methylation, prenylation, oxidation, and reduction to dramatically increase chemical complexity. In
79 plants, the formation of these scaffold molecules and their subsequent decorations are catalyzed by large

enzyme families formed by repeated gene duplication followed by subfunctionalization, neofunctionalization, and gene loss to ultimately produce lineage-specific metabolites. The evolutionary mechanisms that create SM diversity are numerous but include co-option of general metabolism enzymes, evolution of catalytic promiscuity, enzyme compartment switching, the formation of biosynthetic gene clusters, and gene expression changes [6-10]. These evolutionary processes occur across different taxonomic scales, including inter-specific and intra-specific, to generate the chemical variation observed across the plant kingdom.

The Solanaceae, or nightshade family, contains approximately 2700 documented species found on six continents, which collectively have evolved morphological and metabolic adaptations for nearly every environment[11]. A single genus – the *Solanum* – accounts for nearly half of these species[12]. Nightshades grow in environments ranging from deserts to rainforests, with growth habits that vary from epiphytes to trees. The family includes four major food crops (potato, tomato, pepper, and eggplant), a host of minor food crops (including tomatillo, naranjilla, tamarillo, and groundcherry) as well as the several ornamental crops (including petunia, salpiglossis, schizanthus, and brugmansia) and weed species (Jimson weed, and bittersweet). In addition, several Solanaceae species are grown for their narcotic or medicinal properties (tobacco, corkwood tree, deadly nightshade, henbane, and *Datura* species).

The Solanaceae family has become a model system for investigating biodiversity. The Solanaceae community concept was proposed nearly two decades ago, with the idea of using the nightshade family to connect genomics and biodiversity[13]. This concept envisioned harnessing Solanaceae natural diversity for evolutionary studies by creating the necessary network of resources. One important tool was a detailed understanding of Solanaceae phylogenetic relationships (www.solanaceaeesource.org). This framework provides a basis for evolutionary studies within the family. In parallel, the community-driven releases of the first tomato and potato genomes created a genomic foundation. These successful projects spawned numerous additional projects (*e.g.*, SOL-100, Varitome Project, 100 Tomato Genomes Project), resulting in chromosome-scale genome assemblies draft genomes, pan-genomes, resequencing of numerous wild tomato species and cultivars, and an online database for genetic resources[14-20]. As of early 2022, genome sequences are available for more than 30 Solanaceae species (https://plabipd.de/), and it seems likely that many more will follow over the next few years.

These genomic tools are augmented by the availability of comprehensive germplasm resources, particularly for the major crop species of the Solanaceae. These resources allow genetic analysis of phenotypes of interest, facilitate genotype to phenotype comparisons and allow exploration of natural

111 phenotypic diversity. The pioneering work of Charles Rick – and creation of seed stock centers (*e.g.,* GRIN-
112 Global and C.M. Rick Tomato Genetics Resource Center) provide access to crop and wild relative
113 germplasm. Notably, connecting genotype to phenotype within tomato has been greatly accelerated by
114 the development of the introgression lines (ILs) and backcrossed introgression lines (BILs) of wild tomato
115 *S. pennellii* within a cultivated tomato background[21, 22]. These ILs and BILs were instrumental in discovering
116 genes underlying multiple phenotypes, including those related to metabolism[22-25]. In addition, the ability
117 to perform RNA interference (RNAi), virus-induced gene silencing (VIGS), and CRISPR/Cas9 tools in
118 multiple Solanaceae species allows the functional characterization of candidate gene and a more precise
119 connection of genotype and phenotype[26-29].

120 The Solanaceae has emerged as a model system for investigating the biosynthesis and evolution of
121 specialized metabolism (Figure 1). Members of the family have evolved to synthesize several classes of
122 bioactive and lineage-specific specialized metabolites, including phenylpropanoids, acylsugars, terpenes
123 and distinct groups of alkaloids (Figure 2). These specialized metabolites are of interest because they
124 influence fruit aroma and quality and are of potential use as biopesticides and pharmaceuticals. The
125 development of genomic resources, coupled with the ability to survey metabolite variation across diverse
126 germplasm, and to place the resulting data within a phylogenetic context, enabled elucidation of the
127 biosynthesis and evolutionary trajectories of several major classes of Solanaceae SMs.

128

129 **2. Fruity: GWAS-enabled discovery of aroma variation during ripening.**

130 The ripening of fleshy fruits is an agriculturally- and ecologically- important developmental process that
131 makes fruits palatable and facilitates seed dispersal. Although fleshy fruits are highly diverse in
132 morphology and flavor, ripening generally involves cell wall disassembly and associated softening, the
133 conversion of starch into sugars, changes in color, and the biosynthesis of aroma volatiles. Fruit flavor and
134 aroma is a complex species-specific quantitative trait involving the interaction between GM pathways,
135 such as those influencing the accumulation of sugars and organic acids, as well as multiple SM pathways
136 that yield aroma volatiles[30]. Tomato is the long-standing model crop species for investigating ripening
137 mechanisms, including flavor and aroma biosynthesis.

138

139 Recent progress in understanding the genetic and biochemical basis of tomato flavor was facilitated by
140 large-scale genome sequencing and resequencing projects involving hundreds of phenotypically diverse
141 cultivated tomato accessions and wild relatives. These studies revealed insights into the nature of the

142 tomato pan-genome and sequence variation associated with crop domestication and improvement,
143 including gene duplication, single nucleotide polymorphisms, insertion-deletions, and large-scale
144 structural variants [16, 17, 30, 31]. The development of these resources facilitates the identification of genetic
145 variation underlying phenotypic traits via genome-wide association studies. Notably, this approach was
146 successfully deployed for the identification of genetic components underlying variation in tomato fruit
147 flavor and aroma, revealing how human selection for visible traits such as fruit size, yield, and color can
148 lead to alternative outcomes and unintentionally influence SM pathways that contribute to fruit quality.
149
150 Several hundred volatiles are detectable in ripening tomato fruits, but consumer taste panels identified
151 33 metabolites associated with consumer liking and 37 correlated with flavor intensity [30]. These influential
152 aroma volatiles are derived through diversion of general metabolites, including carotenoids,
153 phenylalanine, isoleucine/leucine, and fatty acids into diverse SM pathways. Genetic variation is evident
154 across tomato varieties and 13 fruit aroma volatiles are significantly reduced in a collection of 48 modern
155 cultivars when compared to 236 heirloom tomato varieties. This work shows that breeding of modern
156 varieties for traits such as yield, shelf-life, and disease resistance has inadvertently and negatively altered
157 SM pathways that produce aroma volatiles associated with consumer preference[30]. Subsequent GWAS
158 analyses performed using a panel of 398 diverse tomato accessions analyzed for 27 volatiles along with
159 glucose, fructose, malic acid, and citric acid revealed the existence of 251 association signals for 20 traits,
160 including 15 correlated with aroma volatile production.
161
162 Among these associations are five loci that influence the production of carotenoid-derived volatiles. Two
163 loci specifically influence the production of geranylacetone, which is formed by oxidative cleavage of the
164 minor tomato fruit carotenoids phytoene, phytofluene, ζ-carotene, and neurosporene. A single locus
165 specifically influences 6-methyl-5-hepten-2one (MHO) accumulation, which is derived from lycopene, the
166 main carotenoid pigment in red-fruited tomato varieties. Two additional loci are associated with the
167 production of both geranylacetone and MHO. Analysis of allele frequencies at these loci indicate that
168 genetic complexity was progressively lost during breeding to the point where essentially only two allele
169 combinations associated with accumulation of both volatiles persist in most modern cultivars. Analysis of
170 MHO levels in genotypes with distinct allele combinations revealed that, as breeders selected for high
171 lycopene in red-fruited varieties, they inadvertently selected favorable alleles that increase MHO
172 production. In contrast, the favorable alleles that promote geranylacetone accumulation are absent in
173 modern cultivars[30].

174

175 GWAS also revealed the identity of loci important for producing lipid and phenylalanine-derived volatiles.
176 Ripening tomato fruit accumulate C5 and C6 volatiles derived from the breakdown of linolenic and linoleic
177 acid, which are released from glycerolipids such as triacylglycerol. GWAS analyses of the panel of 398
178 tomato accessions described above identified a chromosome 9-localized SNP that is significantly
179 associated with the fatty acid derived volatiles *Z*-3- hexen-1-ol and hexyl alcohol[32]. This SNP lies within a
180 metabolic QTL region known to influence lipid content in tomato fruit[33]. *Solyc09g091050* (*Sl–LIP8*) was
181 identified as a candidate gene close to this SNP and gene expression analysis revealed that accessions
182 possessing the reference allele from the Heinz 1706 variety had increased levels of *Z*-3-hexen-1-ol and
183 hexyl alcohol together with elevated *Soly09g091050* transcripts. Confirmation that *Sl–LIP8* is responsible
184 for lipid-derived volatile synthesis was achieved through CRISPR/Cas9 gene editing and *in vitro*
185 biochemical assays. The knock-out mutants showed reductions in two C5 (1-pentanol and 1-penten-3-ol)
186 and three C6 (*Z*-3-hexen-1-ol, *E*-2-hexen-1-ol, and hexyl alcohol) volatiles, while the recombinant enzyme
187 catalyzed release of fatty acids from various glycerolipids[32]. The resultant free fatty acids undergo
188 peroxidation at either the C9 or C13 positions in reactions catalyzed by 9-lipoxygenases and 13-
189 lipoxygenases, respectively to yield aroma volatiles.

190

191 The phenylalanine-derived volatiles guaiacol, eugenol, and methylsalicylate contribute to the aroma of
192 tomato fruits and are associated with smoky and medicinal-like aromas, which are often negatively
193 correlated with consumer liking[34]. Guaiacol, eugenol, and methylsalicylate accumulate in tomato fruits as
194 diglycosides, and cleavage of the glycoside groups leads to release of the volatiles in "smoky" cultivars. In
195 contrast, in "non-smoky" varieties these metabolites exist as non-cleavable triglycosides resulting in
196 reduced levels of volatile release[35]. Formation of guaiacol, eugenol, and methylsalicylate triglycosides
197 from their diglycoside precursors is catalyzed by the UDP-glucosyltransferase enzyme, NON-SMOKY
198 GLYCOSYLTRANSFERASE1 (NSGT1). The *NSGT1* gene resides at a locus on chromosome 9 that contains a
199 second gene designated *NSGT2*. Both genes contain structural changes in "smoky" cultivars that are
200 predicted to render them non-functional although the exact structure of the locus was unresolved[35].

201

202 The recent development of 14 new reference tomato genomes assembled using Oxford Nanopore long
203 read sequencing technology allowed the genome structure flanking the *NSGT1* locus to be resolved. Five
204 haplotypes were identified revealing evidence of intraspecific gene duplication and loss at an SM locus
205 that was selected during crop improvement[17]. Haplotype I is proposed to be ancestral and contains

206 predicted functional copies of *NSGT1* and *NSGT2*. All other haplotypes contain coding sequence mutations
207 in *NSGT2.* In addition, haplotypes IV and V also lack functional copies of *NSGT1* and are therefore null
208 mutations for both *NSGT1* and *NSGT2*. Analysis of guaiacol levels across two GWAS panels and within an
209 F_2 population segregating for haplotype V and a functional copy of *NSGT1* demonstrated that fruit guaiacol
210 levels are reduced in individuals that contain a functional copy of *NSGT1*. Together, these data illustrate
211 the combined power of genome sequences developed using long-read sequencing data and GWAS to
212 investigate the evolution of loci associated with SM phenotypes, particularly when the variation is
213 mediated by tandem gene duplication that may be unresolved in genome assemblies derived from short-
214 read data. Overall, these studies represent an example of fundamental science that provides
215 opportunities to breed tomato varieties with favorable aroma volatile alleles.
216

217 **3. Sticky: Single-cell biochemical genetics reveals acylsugar metabolic complexity**

218 Acylsugars are specialized metabolites produced in numerous plant families including the Solanaceae,
219 Convolvulaceae, Geraniaceae, Martyniaceae, Rosaceae, Brassicaceae, and Caryophyllaceae[36-45]. Many
220 species across the Solanaceae produce acylsugars in hair-like Type I- and IV-glandular trichomes, while
221 some species are documented to accumulate acylsugars in fruit pericarp or root exudates[36, 46-48].
222 Acylsugars are composed of a sugar core, most commonly sucrose, and various fatty acids esterified to
223 the core (Figure 3). Despite these simple components, variations in acylation position, chain length, chain
224 branching pattern, and sugar core can result in hundreds of chromatographically separable acylsugars in
225 a single species[37]. Solanaceae acylsugars are the most extensively characterized acylsugar type with more
226 than 100 distinct NMR-resolved chemical structures[36, 49-57]. Acylsugars defend against microbes and
227 insects; for example, deterring whitefly oviposition[58], aphid settling[59], fungal growth[60], and mediating an
228 ant-hornworm-tobacco interaction[61].
229

230 **3.1. Harnessing acylsugar genotypic diversity for tomato pathway determination**

231 Tomato acylsugar diversity was employed to uncover the acylsugar biosynthesis pathway within cultivated
232 tomato, *S. lycopersicum*. Analysis of *S. lycopersicum* introgression lines carrying *S. pennellii* chromosomal
233 segments was instrumental in identifying loci required for acylsugar biosynthesis[24, 62]. The identification
234 and subsequent validation of candidate genes was facilitated by trichome-specific transcriptome, *in vitro*
235 enzyme assays, and *in vivo* gene VIGS knockdown and CRISPR/Cas9 knockout. These approaches
236 uncovered the core acylsugar pathway in *S. lycopersicum* glandular trichomes. A series of evolutionarily
237 related BAHD acyltransferases, named AcylSucrose AcylTransferase 1-4 (ASAT1-4), acylate sucrose

238 sequentially to produce tetraacylsucroses consisting of acyl chains at R_2, R_3, R_4, and $R_{3'}$[24, 63, 64](Figure 4).
239 Each enzyme selectively acylates specific sucrose hydroxyls with varying promiscuity for acyl-CoA
240 substrates. Documenting this pathway enabled discovery of mechanisms responsible for acylsugar
241 diversity in wild tomato relatives.
242 Intra- and inter-specific differences in tomato acylsugar structures result in part from differing ASAT
243 activities. Comparative biochemical analysis of cultivated and wild tomato ASAT sequences uncovered
244 amino acid residues responsible for specific activity differences. For example, the comparison of ASAT2
245 sequences and *in vitro* enzyme activities across tomato species revealed two mutations that impact acyl-
246 CoA specificity. Residues Val/Phe[408] and Ile/Leu[44] influence the ability to use the structurally similar iC5-
247 CoA and aiC5-CoA, respectively, without altering activity with nC12-CoA[64]. Comparison of *S. lycopersicum*
248 and *S. habrochaites* ASAT3 homologs revealed a Tyr/Cys[41] residue change impacting the enzyme's ability
249 to use nC12-CoA[63]. Characterization of *S. habrochaites* ASAT4 in accessions collected from Ecuador to
250 Southern Peru revealed variations in acetylation patterns that were explained either by changes in ASAT4
251 expression or coding sequence mutations[65, 66]. The comparative biochemistry approach revealed
252 differences in enzyme acyl donor specificity, which impacted acylsugar phenotypes. This approach also
253 determined evolutionary changes in enzyme acyl acceptor specificity.

254 *S. pennellii* LA0716 produces acylsucroses through a 'flipped pathway', resulting from changes in ASAT
255 acyl acceptor specificity[67]. While cultivated tomato produces acylsucroses with one furanose ring
256 acylation (termed F-type acylsucroses), *S. pennellii* and some *S. habrochaites* accessions synthesize
257 acylsucroses acylated exclusively on the pyranose ring[63]. These 'P-type' acylsucroses are produced by
258 alternate ASAT2 and ASAT3 homologs, which catalyze the third and second pathway steps, respectively.
259 The published results suggest that *S. pennellii* ASAT2 likely evolved from an ancestral enzyme capable of
260 acylating both mono- and diacylsucrose. Analogous sequence changes in ASAT3, potentiated by ASAT3
261 duplication, resulted in the neofunctionalized ASAT3 duplicate found in *S. habrochaites* and *S. pennellii*.
262 This study revealed a remarkably small number of amino acid changes that caused a major change in
263 pathway structure and product phenotypes in closely related species.

264 The flipped *S. pennellii* pathway and recruitment of an invertase-like enzyme appear to have potentiated
265 evolution of *S. pennellii* acylglucose synthesis (Figure 4). *S. pennellii* acylglucoses are synthesized from P-
266 type acylsucroses by a neofunctionalized glycoside hydrolase 32 family (GH32) beta-fructofuranosidase,
267 SpASFF1[68]. The modified SpASFF1 substrate binding site correlates with a derived P-type acylsucrose
268 cleavage activity, yet the neofunctionalized enzyme does not act on the F-type acylsucrose produced by

269 *S. lycopersicum*. In addition, SpASFF1 lacks activity with sucrose, associated with changes to the canonical
270 sucrose binding pocket. Instead, the modified SpASFF1 substrate binding site correlates with a derived P-
271 type acylsucrose cleavage activity, yet the neofunctionalized enzyme does not act on the F-type
272 acylsucrose produced by *S. lycopersicum*. SpASFF1 specificity for P-type acylsucroses supports the
273 hypothesis that P-type acylsucroses are required for acylglucose production. Indeed, cultivated tomato
274 lines engineered to contain both the flipped pathway and SpASFF1 accumulate acylglucoses. This indicates
275 that acylglucose biosynthesis requires both a neofunctionalized invertase and the *S. pennellii* flipped
276 pathway. Finally, CRISPR/Cas9 deletion of SpASFF1 led to accumulation of only acylsucroses – without
277 detectable acylglucoses – in *S. pennellii*, reinforcing that the neofunctionalized invertase is necessary for
278 acylglucose synthesis in the wild tomato. SpASFF1 invertase is an example of co-option of general
279 metabolic enzyme to specialized metabolism into acylsugar biosynthesis – in this case resulting in different
280 sugar core composition.

281 The theme of GM enzymes recruitment to SM by gene duplication, changes in gene expression and
282 enzyme structure and function also contribute to acyl chain type variation. For example, the duplicated
283 and neofunctionalized isopropylmalate synthase gene, IPMS3, influences isoC5 acyl chain abundance[69]. In
284 contrast to the canonical Leu biosynthetic IPMS, IPMS3 expression is restricted to type I/IV glandular
285 trichome tip cells, and the *S. lycopersicum* enzyme is insensitive to Leu-mediated feedback inhibition *in*
286 *vitro* due to truncation of the C-terminal allosteric regulatory domain. Apparently, the lack of this domain
287 frees the enzyme from Leu feedback regulation, enabling pathway diversion. IPMS3 allelic variation
288 directly correlated with abundance of isoC5 and isoC4 acyl chains in wild *S. pennellii* accession acylsugars;
289 accessions with majority isoC4 acyl chains were homozygous for a truncated, inactive IPMS3. In contrast,
290 isoC5 acyl chains were abundant in accessions either heterozygous or homozygous for the unregulated
291 IPMS3. These results reveal that acyl-CoA availability influences acylsugar acyl chain composition.

292 Further evidence for this hypothesis was provided by identification of natural chain diversity associated
293 with allelic diversity of two acyl-CoA biosynthesis genes[70]. These trichome-expressed genes, an enoyl-CoA
294 hydratase (AECH1) and acyl-CoA synthetase (AACS1), reside in a gene cluster syntenic to the chromosomal
295 region containing ASAT1. The Solanaceae family shares the syntenic region, which was likely derived from
296 a Solanaceae-specific polyploidy event. Silencing AECH1 and AACS1 in *S. lycopersicum*, *S. pennellii*, and
297 the more distantly related *Solanum quitoense,* reduced or eliminated medium length (10-12 carbons) acyl
298 chains from acylsugars. Additionally, the presence of AECH1 and AACS1 correlates with natural variation
299 in medium acyl chains. For example, in the short chain producing genera *Petunia* and *Nicotiana*, AECH1

300 and AACS1 are either missing or present as pseudogenes. These genes represent another example of how
301 evolutionary changes in metabolic machinery impacted acylsugar composition.

302 **3.2. Genomics tools enable comparative biochemistry in non-model organisms**

303 Application of DNA sequencing, modern analytical chemistry, and reverse genetic tools such as VIGS and
304 genome editing enabled documentation of additional acylsugar evolutionary mechanisms in non-model
305 species. LC-MS screening and NMR-resolved structural analysis identified Solanaceae species that produce
306 unique acylsugars with varying cores, acylation positions, and chain types[37, 50, 53, 57, 71, 72]. For example,
307 extant members of early-diverging lineages produce acylsucroses with acylation patterns undocumented
308 in cultivated and wild tomatoes. Additionally, acylated glucoses are detected in some species within the
309 *Petunia*, *Nicotiana*, *Datura*, and *Solanum* genera[72-76]. Within the large *Solanum* genus, *myo*-inositol sugar
310 cores have been documented in *S. lanceolatum, S. quitoense*, and *S.* nigrum[71, 72, 77]. Evolution of acylsugar
311 biosynthesis was investigated in four non-model species: *Salpiglossis sinuata*, *Petunia axillaris*, *S. nigrum*,
312 and *S. quitoense*. Comparison of the enzymes and pathways in each species revealed features of long-
313 term and clade-specific acylsugar traits.

314 ***3.2.1. Inferring early events in acylsugar evolution***

315 Investigations of two members of early diverging lineages, *S. sinuata* and *P. axillaris*, revealed acylsugar
316 biosynthesis evolutionary changes occurring over tens of millions of years (My), well beyond the
317 approximately 7 My of Solanum tomato clade history[11, 37, 78]. Despite similarity of acylation positions
318 between tomato species, *S. sinuata* and *Petunia* acylsugars, a major shift occurred in the acylsugar
319 biosynthetic pathway. The ancestral pathway found in *S. sinuata* and *P. axillaris* begins with a sucrose-
320 acylating ancestral ASAT1, aASAT1, which is not found in tomato clade species. Another surprise is that
321 the SlASAT1 and SlASAT2 orthologs, aASAT2 and aASAT3, respectively catalyze the second and third
322 acylations. The first three acylations by the early evolving aASAT1-3 pathway produce triacylsucroses with
323 the same three positions acylated as SlASAT1-3. Coinciding with this, aASAT2 and aASAT3 retained their
324 selectivity for the R4 and R3 of sucrose, respectively, but shifted acyl acceptor specificity to free and
325 monoacylsucrose, respectively. This activity shift correlates with aASAT1 loss in species with modern
326 acylsugar biosynthesis pathways. Transcriptome and genome analyses suggest that the aASAT1 gene
327 disappeared from the last common ancestor of the *Capsicum* and *Solanum* genera, ~15-20 MYA.
328 Identification of these ancestral acylsugar pathways support sucrose as the ancestral acyl acceptor. From
329 these studies of early-diverging Solanaceae species, ASAT gene loss and neofunctionalizations were
330 implicated in a changing acylsucrose pathway, analogous to those described above in the case of the *S.*
331 *pennellii* flipped acylsucrose pathway.

332

333 The ancestral and derived acylsucrose pathways provide insight into the evolutionary origins of
334 acylsugars[37]. Lamiidae BAHD sequence homology, phylogenetics, and known whole genome duplication
335 events all enabled inferences regarding early acylsugar evolution. One hypothesis, based on sequence
336 analysis, is that ASAT sequences derive from an alkaloid biosynthetic enzyme ancestor. Based on
337 nonsynonymous mutation rates and historical polyploidy events, the clade containing ASAT1,2,3 appears
338 to have arisen via an ancient whole genome duplication before the Solanaceae-Convolvulaceae split (~50-
339 65 MYA). Subsequent duplications prior to, and following the Solanaceae polyploidization, led to evolution
340 of the ASATs and paralogs found in the ASAT1,2,3 clade. As described above, our model of acylsugar
341 biosynthetic pathway evolution invokes loss of aASAT1, refinement of ASAT1 and ASAT2 activities, and
342 recruitment of ASAT3 occurred later in Solanaceae diversification.

343

344

345

346 ***3.2.2. Acylhexoses in non-model plants***

347 Metabolite profiling revealed that, like *S. pennellii*, black nightshade (*Solanum nigrum*) also produces
348 acylglucoses, an observation that enabled discovery of convergent and new acylsugar enzyme activities.
349 *S. nigrum* creates di- and triacylglucoses through a similar, yet distinct, pathway when compared to *S.*
350 *pennellii* acylglucose biosynthesis[72](Figure 4). Both pathways proceed through a series of sucrose
351 acylations, followed by action of an acylsugar fructofuranosidase. The *S. nigrum* invertase, SnASFF1, and
352 SpASFF1 enzymes share similarities including a modified DDTK sucrose binding pocket, loss of canonical
353 invertase activity cleaving sucrose, and neofunctionalized activity with acylsucroses. However, each ASFF1
354 enzyme resides in a distinct glycoside hydrolase subfamily 32 clade and cleaves different substrates:
355 triacylsucroses by SpASFF1 and diacylsucroses by SnASFF1. SnAcylGlucoseAcetylTransferase1, SnAGAT1,
356 catalyzes the third *S. nigrum* acylation, marking yet another distinction between *S. nigrum* and *S. pennellii*
357 triacylglucose biosynthesis; this is the only enzyme to acylate an acylglucose described to date. As the two
358 characterized *Solanum* acylglucose biosynthetic pathways include distinct invertases, it is plausible that
359 this mechanism evolved in other acylglucose-producing genera.

360 In contrast to the detailed information available for acylsucrose and acylglucose biosynthesis, the pathway
361 leading to acylinositol synthesis in the *Solanum* remains largely enigmatic. So far only one enzyme was
362 demonstrated in acylinositol biosynthesis: the *S. quitoense* enzyme TriAcylInositol AcetylTransferase,
363 SqTAIAT, acetylates triacylinositols to produce tetraacylinositols[71]. SqTAIAT is the closest known *S.*

364 *quitoense* homolog to the final enzyme in tomato acylsucrose biosynthesis, SlASAT4, indicating
365 conservation of acetyltransferases across acylinositol and acylsucrose biosynthesis. Both enzymes
366 acetylate triacylsugars differing in their sugar core. Similar enzymatic activity and high sequence similarity
367 suggest a common evolutionary origin for acylinositol and acylsucrose biosynthesis. However, the initial
368 steps of acylinositol biosynthesis remain unresolved. Further pathway elucidation in *S. quitoense* and *S.*
369 *nigrum* may uncover the evolutionary innovations underlying acylinositol production.
370

371 **3.3. Into the depths with acylsugars**

372 It was recently shown that cultivated tomato accumulates acylsugars in roots and root exudates[48]. Tomato
373 root acylsugars structurally differ from those in trichomes, contrasting in acyl chain type, acyl chain
374 number, and sugar core type. For example, six- and seven-carbon acyl chains and glucose sugar cores are
375 only detected in the roots. These structural differences suggest evolutionary changes in the underlying
376 biochemistry. One key observation is that characterized tomato trichome-expressed ASAT transcripts
377 were not detected in root tissue, although they do express closely related homologs. These expression
378 data suggest the hypothesis that roots produce acylsugars through an alternative pathway. In fact,
379 expression of two ASAT4 paralogs correlates with acylsugar abundance in roots. While the function of
380 root acylsugars is unknown, different microbial communities systemically impacted root exudate
381 acylsugar abundances[48]. Investigating root acylsugar metabolism may unearth a root-specific acylsugar
382 biosynthetic pathway among other tantalizing prospects.
383

384 **4. Stinky: Variations on a theme define terpene diversity across *Solanum***

385 Terpenoids are structurally diverse and are produced across all kingdoms of life, yet all are derived from
386 the simple five-carbon isomers, dimethylallyl diphosphate (DMAPP) and isopentenyl diphosphate (IPP).
387 These precursors are formed through either the mevalonate (MVA) or 2-C-methyl-D-erythritol 4-
388 phosphate (MEP) pathways[79]. Plants are unique in that they contain both the cytosolic MVA pathways and
389 the plastid localized MEP pathway; having evolved to generate substantial flux towards DMAPP and IPP
390 as well as create separate subcellular pools of these metabolites for different pathways[79]. Terpenoids have
391 diverse functions ranging from the production of photosynthetic pigments and ubiquinone in the electron
392 transport chain to the production of several classes of plant hormones. However, most plant terpenoids
393 are lineage-specific specialized metabolites with C10 – C30 carbon skeletons that provide a fitness benefit
394 to the host organism through signaling and defense[79].
395

396 Plant terpenoid diversity is created at multiple levels. Firstly, small gene families produce *cis* and *trans*-
397 prenyltransferases that initially condense a single molecule of DMAPP and IPP to form either geranyl
398 diphosphate (GPP) (*trans* isomer) or neryl diphosphate (NPP) (*cis* isomer). These C10 metabolites can then
399 be extended by five carbon units, through condensation with additional units of IPP, to yield *trans*- or *cis*-
400 farnesyl diphosphate (*E*,*E*-FPP or *Z*,*Z*-FPP, C15), geranylgeranyl or nerylneryl diphosphate (GGPP or NNPP,
401 C20), or longer chain prenyl diphosphates[79]. Short-chain prenyl diphosphates (C10-C20) are substrates for
402 terpene synthases (TPS), which exist as moderately large gene families (up to ~100 members) and catalyze
403 the formation of hydrocarbon terpene skeletons via rearrangements and cyclization. TPS enzymes possess
404 considerable catalytic potential. They frequently utilize more than one substrate, and catalysis by a single
405 enzyme often generates multiple products[79-81]. These hydrocarbon terpene skeletons are often
406 functionalized by the addition of hydroxyl groups, which provide targets for modifications such as
407 epoxidation, methylation, acylation, and glycosylation, ultimately generating the vast complexity of
408 terpenoids observed across the plant kingdom.
409
410 The availability of a high-quality reference genome assembly for cultivated tomato (*Solanum*
411 *lycopersicum*) facilitated what is likely the most comprehensive published catalogue of terpene scaffold
412 biosynthesis in plants. The data highlight considerable chemical complexity with *in vitro* biochemical data
413 revealing the potential to synthesize 53 known hydrocarbon terpene scaffolds plus several unidentified
414 products. These terpenes arise through combined catalysis of seven *cis*-prenyltransferases and 10 *trans*-
415 prenyltransferases that form C10, C15, and C20 prenyl diphosphates, together with 34 functional TPS
416 enzymes[82, 83]. Consistent with the known catalytic promiscuity of TPS enzymes, many of the tomato TPSs
417 can utilize more than one substrate, particularly the sesquiterpene synthases that use both *E*,*E*-FPP and
418 *Z*,*Z*-FPP, and yield multiple products. In addition, considerable catalytic redundancy exists. For example,
419 eight distinct TPSs catalyze the formation of the monoterpene β-myrcene. Individual CPT, TPT, and TPS
420 enzymes are localized to the cytosol, plastids, as well as mitochondria, and the corresponding genes are
421 differentially expressed across tomato tissues: this highlights the spatial separation of terpene synthesis
422 modules across tomato. Metabolite profiling of 13 tomato tissues identified 29 out of 53 terpenes *in*
423 *planta*, suggesting that some terpenes are either below the limit of detection in tomato grown under
424 standard cultural conditions or are further modified to produce more structurally complex metabolites.
425
426 Genomic clustering is a key feature of terpene biosynthetic genes in plants[84]. These clusters generally
427 consist of both paralogs and non-homologous genes encoding enzymes of terpene biosynthesis, creating

428 a reservoir for the evolution of chemical novelty and facilitating the inheritance of SM modules that
429 promote plant adaptation. Gene duplication within these clusters is often followed by pseudogenization
430 and gene loss to create additional chemical variation. The majority of the 52 TPS loci in tomato, including
431 18 predicted pseudogenes, are located within gene clusters dispersed across the genome[82]. In addition,
432 the *TPS* gene clusters on chromosomes 6, 7, 8, and 12 also contain combinations of *cis* or *trans*
433 prenyltransferases, cytochromes P450, methyltransferases, acyltransferases, and glycosyltransferases[82,
434 85]. While most of the potential terpene modifying enzymes within these clusters await functional
435 characterization, a three-gene subcluster on chromosome 8 comprising *SlTPS21-CYP71D51-SlCPT2* was
436 demonstrated to synthesize (+)-lycosantalonol from NNPP[86].
437
438 Along with the existence of the 18 *TPS* pseudogenes in the tomato genome, three *TPS*-related gene
439 clusters on chromosomes 6, 8, and 12 also contain inactive cytochromes P450 genes[82]. The high
440 prevalence of pseudogenes within these tomato terpene biosynthetic gene clusters suggests that there is
441 potential for considerable genetic variation. For example, a gene that is pseudogenized in one accession
442 or species may be functional in another. Thus, variation in terpene-related gene clusters may exist
443 between distinct accessions of *S. lycopersicum* but also more likely across the genomes of diverse
444 Solanaceae species. The increasing availability of high-quality chromosome scale reference genomes
445 assembled from long-read sequencing will facilitate identification of additional gene clusters and future
446 comparative evolutionary analysis of terpene biosynthesis across the Solanaceae.
447
448 Within the *Solanum* genus, distinct evolutionary trajectories associated with trichome-derived terpene-
449 related gene clusters are indeed apparent between cultivated tomato and wild relatives that diverged
450 from a common ancestor approximately two-three million years ago[11]. Notably, while limited terpene
451 diversity exists in trichomes between cultivated tomato accessions, considerable variation is observed
452 across distinct populations of *Solanum habrochaites* and between *S. habrochaites* and *S. lycopersicum*[87].
453 This genetic variation determines whether specific accessions preferentially synthesize monoterpenes
454 (C10) or sesquiterpenes (C15), and results from differences at the *cis-prenyltransferase 1* (*CPT1*) locus and
455 associated TPS-e/f enzymes that are located within the chromosome 8 terpene gene cluster[85]. For
456 example, trichomes of cultivated tomato predominantly accumulate the monoterpene β-phellandrene,
457 which is synthesized from NPP by neryl diphosphate synthase1 (NDPS1)[88]. While select monoterpene-
458 producing accessions of *S. habrochaites* also contain an ortholog of NDPS1, a separate group of
459 sesquiterpene producing accessions of *S. habrochaites* possess the C15-producing *Z,Z*-farnesyl

460 diphosphate synthase (zFPS) at the *CPT1* locus[89, 90] (Figure 5). Comparative sequence analysis, homology
461 modeling, and site-directed mutagenesis revealed that the relative positioning of bulky aromatic amino
462 acid residues within a hydrophobic cleft specifies substrate binding and prenyl-chain elongation between
463 CPT1 isoforms with NDPS1 and zFPS activity and that this contributes to intraspecific terpene variation in
464 *S. habrochaites*[90].
465
466 Together with divergent CPT1 enzymes, terpene diversity in *S. habrochaites* trichomes is also driven by
467 natural variation in chromosome 8 cluster TPS-e/f subfamily members. *S. lycopersicum*, synthesizes a
468 cocktail of monoterpenes in trichomes from NPP using the TPS-e/f enzyme, β-phellandrene synthase
469 (SlPHS1 / SlTPS20)[88]. PHS1 activity is conserved in some *S. habrochaites* accessions while others contain
470 the TPS-e/f paralogs limonene synthase (ShLMS) and pinene synthase (ShPIS), which catalyze the
471 formation of limonene and α-pinene from NPP, respectively[87]. In addition to this intraspecific variation in
472 monoterpene biosynthesis, two additional groups of *S. habrochaites* accessions possess TPS-e/f enzymes
473 that synthesize sesquiterpenes from *Z*,*Z*-FPP produced by zFPS: santalene and bergamotene synthase
474 (ShSBS) catalyzes the formation of a mixture of santalene and bergamotene isomers[87, 89]. In contrast, a
475 distinct, yet closely related enzyme, zingiberene synthase (ShZIS) catalyzes the formation of 7-
476 epizingiberene[87] (Figure 5). These sesquiterpene forming TPS-e/f enzymes are not present in *S.*
477 *lycopersicum* and, to date, appear to be restricted to a subset of *S. habrochaites* accessions. Overall,
478 together with variation at the *CPT1* locus, these examples illustrate the evolutionary potential of SM
479 associated gene clusters to create and maintain inter-specific and intra-specific chemical diversity. This
480 relatively rapid intra-specific evolution of chemical variation in specific populations of plants may confer
481 selective advantage against diverse biotic challenges.
482
483 The ability of trichomes of select *S. habrochaites* accessions to synthesize the sesquiterpenes santalene
484 and bergamotene as well as 7-epizingiberene and their derivatives is known to confer increased tolerance
485 to insect pests and pathogens when compared to trichomes that synthesize *S. lycopersicum* type
486 monoterpenes[91-94]. Santalene and bergamotene backbones are oxidized into sesquiterpene acids via
487 unknown enzymes[93]. In contrast, 7-epizingiberene is sequentially oxidized to a combination of 9-hydroxy-
488 zingiberene and 9-hydroxy-10,11-epoxy-zingiberene in reactions catalyzed by the trichome-expressed
489 cytochrome P450, ShCYP71D184[95] (Figure 5). 9-hydroxy-10,11-epoxy-zingiberene is particularly effective
490 in bioactivity assays against whiteflies (*Bemisia tabaci*) and the microbial pathogens, *Phytophthora*
491 *infestans* and *Botrytis cinerea*. ShCYP71D184 is encoded by the *Sohab01g008670* locus and is therefore

492 not located in the chromosome 8 TPS cluster responsible for the synthesis of the 7-epizingiberene
493 substrate. The predicted ShCYP71D184 protein is 94% identical to its putative ortholog from *S.*
494 *lycopersicum* SlCYP71D184 / Solyc01g008670. The function of SlCYP71D184 is unknown but *S.*
495 *lycopersicum* trichomes do not synthesize 7-epizingiberene and this enzyme is incapable of catalyzing the
496 formation of 9-hydroxy-zingiberene and 9-hydroxy-10,11-epoxy-zingiberene. Although not completely
497 understood, these data suggest that, like other loci that influence terpene biosynthesis in glandular
498 trichomes of *Solanum*, genetic variation exists at the *CYP71D184* locus that specifies chemical diversity.
499
500 **5. Spicy: Lineage-specific biosynthesis of capsaicinoids in pepper.**
501 Species within the *Capsicum* genus of the Solanaceae possess the capacity to synthesize a group of
502 specialized metabolites known as capsaicinoids, including capsaicin, the principal determinant of
503 pungency in chili peppers. These specialized metabolites are of culinary and cultural importance but also
504 possess applications as topical pain medications and show efficacy as anti-inflammatories, treatments for
505 cancer and weight-loss, and possess anti-microbial activities[96-99]. Capsaicinoids are synthesized within the
506 placenta that surrounds the seeds of developing fruit and act as feeding deterrents for small mammals
507 such as rodents, but not birds[100]. This deterrence is mediated by the mammalian vanilloid receptor 1 (VR1)
508 ion channel that is localized to sensory nerve endings and responds to heat stimuli[101]. The ortholog of VR1
509 from birds does not respond to capsaicin and as such, birds, which are more efficient seed dispersers than
510 small mammals, are unaffected by the pungency of pepper fruits[102].
511
512 The biosynthesis of capsaicinoids is not fully understood, particularly at the biochemical level and this
513 pathway is yet to be reconstructed in a heterologous system. However, capsaicin biosynthesis is
514 considered a derived trait within *Capsicum*, as species from the more ancient Andean clade of the genus
515 are non-pungent[103]. Within *Capsicum* species, intra-specific variation exists resulting in loss of
516 pungency[103]. Most notably, this intra-specific variation occurs in the major crop species *Capsicum annuum*
517 and gives rise to both pungent and sweet pepper cultivars[103]. Capsaicin is synthesized through the
518 condensation of vanillylamine, derived from the phenylpropanoid pathway, with 8-methyl-6-nonenoyl-
519 CoA, produced through branched-chain amino acid metabolism and fatty acid synthesis[104]. Genetic
520 analyses identified loci associated with capsaicin accumulation and genes within the phenylpropanoid,
521 branched-chain amino acid catabolism, and fatty acid synthesis pathways are among the candidates
522 discovered[105-107]. For example, loss of function alleles at the *AMT* locus, which encodes an
523 aminotransferase that catalyzes the formation of vanillylamine from vanillin, disrupts capsaicin

524 biosynthesis[108-110]. Similarly, mutation in a ketoacyl-ACP reductase (*CaKR1*), an enzyme involved in fatty
525 acid biosynthesis, resulted in undetectable levels of capsaicin and 8-methyl-6-nonenoic acid, a precursor
526 of 8-methyl-6-nonenoyl-CoA[111]. In addition, the BAHD acyltransferase capsaicin synthase, also known as
527 Pun1, is associated with pungency in hot pepper and proposed to catalyze the condensation of
528 vanillylamine with 8-methyl-6-nonenoyl-CoA to form capsaicin[112]. A 2.5 kb deletion allele at this locus is
529 present in non-pungent genotypes, although biochemical evidence supporting a direct role for this
530 enzyme in capsaicin biosynthesis is lacking[112]. Overall, these studies reveal genetic variation across
531 *Capsicum* that has likely arisen due to domestication and selection.
532
533 **6. Bitter: Evolutionary signatures of glycoalkaloid biosynthesis in *Solanum***
534 Steroidal glycoalkaloids (SGAs) are bitter and toxic metabolites that occur in *Solanum* including the crop
535 species tomato, potato, and eggplant. SGAs provide protection against herbivory as well as microbial
536 pathogens and are proposed to function through the disruption of cell membranes and inhibition of
537 cholinesterase activity[113]. In the United States, SGA levels are monitored in potato to maintain levels
538 below an FDA-regulated threshold due to their toxicity[114]. Evolution and domestication shaped SGA
539 diversity in *Solanum*; metabolite profiling and chemical structure elucidation reveal hundreds of SGAs that
540 differ among members of the genus due to gene gain and loss between species[115, 116]. For example, α-
541 tomatine and esculeoside A accumulate in tomato while α-solasonine and α-solamargine are synthesized
542 in eggplant. In contrast, domesticated potato synthesizes α-solanine and α-chaconine, while leptines,
543 SGAs that display efficacy against Colorado potato beetle (CPB), are found in wild potato species (Figure
544 6) [10, 117-120]. SGAs arise from the modification of cholesterol produced from the mevalonate pathway and
545 are characterized by a nitrogen-containing 27-carbon core, which can undergo multiple glycosylations to
546 form steroidal glycoalkaloids[121]. Comparison of genomic sequences between species revealed that several
547 biosynthetic steps of SGA formation in tomato, potato, and eggplant, encoded by *GLYCOALKALOID*
548 *METABOLISM* (*GAME*) genes, are clustered within these genomes [8, 122].
549
550 Formation of plant SGA sterol cores requires diversion of 2,3-oxidosqualene from the mevalonate
551 pathway into cholesterol biosynthesis, and this biosynthetic pathway appears to have evolved from the
552 duplication and divergence of genes involved in phytosterol biosynthesis, which leads to the production
553 of brassinosteroids, an essential class of phytohormones[121]. Cycloartenol synthase (CAS) converts 2,3-
554 oxidosqualene into cycloartenol, and this metabolite is the branch point between cholesterol and
555 phytosterol biosynthesis as it serves as a substrate for both SSR2 (sterol side chain reductase 2) and SMT1

556 (sterol C-24 methyltransferase) to form cycloartanol or 24-methylenecycloartanol, respectively[121].
557 Cholesterol biosynthesis leads to the production of the SGAs and saponins in both glycosylated and
558 aglycone forms[121]. Elucidation of cholesterol biosynthesis in plants revealed five enzymes shared between
559 the cholesterol and phytosterol pathways[121]. Phylogenetic analysis of enzymes specific to cholesterol
560 biosynthesis suggests that *C5-SD2* (sterol C-5(6) desaturase), *7-DR2* (7-dehydrocholesterol reductase),
561 *SMO3* (C-4 sterol methyl oxidase) and *SMO4* likely arose from duplication and divergence of the
562 phytosterol pathway genes, *C5-SD1*, *7-DR1*, *SMO1* and SMO2[121].
563
564 Presence-absence variation of genes involved in the conversion of dehydro-SGAs to dihydro-SGAs
565 contributes to SGA diversity within *Solanum*. The first spirosolosane-type SGA formed, (*22S, 25S*)-spirosol-
566 5-en-3β-ol, contains a $\Delta^{5,6}$ double bond[10]. In tomato, tomatidine is synthesized from a multistep process
567 starting with the oxidation and isomerization of (*22S, 25S*)-spirosol-5-en-3β-ol to tomatid-4-en-3-one by
568 GAME25, and the addition of four sugars (galactose, glucose, glucose, and xylose) to the C-3 position of
569 tomatidine results in the production of tomatine, the major tomato SGA[7, 123, 124]. Lack of a functional
570 *GAME25* is associated with the production of unsaturated SGAs, including α-solamargine, α-solasonine,
571 and malonylsolamargine in *S. melongena* (eggplant) and expression of tomato *GAME25* in eggplant results
572 in the production of saturated SGAs[123]. However, the mechanism underlying a lack of saturated SGA
573 accumulation in domesticated potato is less clear. A putative *GAME25* homolog is present in the genome
574 of domesticated potato, and recombinant expression of the corresponding enzyme revealed the same
575 activity as the tomato enzyme: 3β-hydroxyl group oxidation and isomerization of the double bond from
576 the C-5,6 position. The potato GAME25 enzyme is active with unsaturated spirolosane- and solanidine-
577 type SGAs although the corresponding saturated SGAs do not accumulate in domesticated potato[123].
578 Overexpression of tomato GAME25 in potato hairy root cultures leads to accumulation of demissidine, a
579 saturated solanidine SGA found in wild potato. This suggests that the downstream enzymatic activities
580 involved in the production of saturated SGAs exist in domesticated potato[125]. However, the mechanism
581 leading to the lack of saturated SGAs in domesticated potato remains unclear, and the *in vivo* function of
582 the domesticated potato GAME25 and expression levels of the corresponding gene remain to be
583 determined[123, 125].
584
585 While the initial steps of spirolosane-type SGA formation are conserved between tomato and potato, SGA
586 biosynthesis diverges in potato to produce solanidine-type SGAs[10]. Potato contains two major solanidane-
587 type SGAs, α-solanine and α-chaconine, which differ only in the identity of the C-3 sugar additions;

588 solanine contains galactose with rhamnose and glucose additions while chaconine contains glucose with
589 two rhamnose additions[10]. The 2-oxoglutarate dependent dioxygenase, DPS (Dioxygenase for Potato
590 Solanidane synthesis), catalyzes solanidine ring formation via C-16 hydroxylation[10]. While both eggplant
591 and tomato contain *DPS* homologs and each recombinant enzyme is capable of C-16 hydroxylation of
592 spirolosane-type SGAs, the expression of the corresponding genes is low or undetectable in eggplant and
593 tomato, which likely explains the lack of solanidine-type SGAs in these species[10]. The *DPS* genes are
594 located on chromosome 1 within a syntenic block that is conserved in *Solanum* and contains additional
595 SM-related genes, suggesting that the *DPS* genes evolved prior to speciation[10]. While some wild potato
596 species, such as *Solanum chacoense*, produce leptines, solanidine-type SGAs that are effective at
597 defending against CPB, domesticated potato does not produce these SGAs. Leptine formation requires
598 the hydroxylation of solanidine-type SGAs by GAME32 and the subsequent acetylation by an unknown
599 enzyme. Tomato and domesticated potato lack a functional GAME32 homolog and the corresponding
600 leptine SGAs[117].
601
602 Domestication and selection for non-bitter fruit to aid in seed dispersal influence SGA content in tomato
603 during fruit ripening. The fruit ripening associated biosynthesis of esculeoside A from α-tomatine
604 alleviates the bitter taste associated with SGAs[117]. The hydroxylation of α-tomatine at the C-23 position is
605 the first committed step of fruit ripening associated SGA accumulation (i.e. esculeoside A), and is catalyzed
606 by the 2-ODD enzyme, GAME31[117, 126]. Esculeoside A formation requires an additional hydroxylation,
607 followed by acetylation, and the glycosylation of acetoxy-hydroxytomatine by GAME5[117, 127, 128]. The export
608 of α-tomatine and α-tomatine derivatives out of the vacuole by a nitrate transporter 1/peptide
609 transporter family (NPF) transporter, GORKY (meaning bitter in Russian), is essential for esculeoside A
610 formation[129]. The sequestration of toxic SGAs to the vacuole likely prevents self-toxicity, and this is
611 evidenced by the observation that tomato plants overexpressing GORKY (facilitating SGA export to the
612 cytosol) displayed severe morphological phenotypes[129]. In contrast, fruit from the same overexpression
613 lines did not display signs of self-toxicity suggesting that the conversion of toxic/bitter SGAs to
614 esculeosides prevents self-toxicity[129].
615
616 The synteny of the metabolic gene clusters involved in SGA production among *Solanum* species highlights
617 the common origin of the trait that diverged between species through loss or gain of function of individual
618 genes to create SGA diversity. Several of the genes involved in spirolosane-type SGA formation are found
619 clustered on potato, eggplant, and tomato chromosomes 7 and 12[8, 122]. Tomato possesses two extra genes

620 in these clusters as potato and eggplant lack homologs of GAME17 and 18, two UDP-glucosyltransferases
621 responsible for the consecutive additions of glucose to tomatidine galactoside during α-tomatine
622 biosynthesis in tomato[8]. Current genomic resources show that pepper (*Capsicum annuum*) does not
623 possess the chromosome 12 cluster or putative orthologs of GAME4 and GAME12 found within the
624 cluster, and this absence likely results in the lack of SGAs in *C. annuum*[122]. The *2-ODD* genes involved in
625 solanidine, leptine, and esculeoside SGA biosynthesis are also clustered with additional *2-ODDs* of
626 unknown function[117]. Changes in gene expression (i.e. low expression of *DPS* tomato homolog) or the
627 presence-absence of single genes (i.e. *GAME32* presence in *S. chacoense*) contribute to SGA diversity in
628 *Solanum*.
629
630 **7. Addictive and Deadly: Convergent and divergent evolution shapes nicotine and tropane alkaloid**
631 **metabolism.**
632 Several Solanaceae genera, including *Datura*, *Atropa*, *Hyoscyamus*, *Mandragora*, and *Scopolia* derive
633 medicinal and toxic qualities from the biosynthesis of tropane alkaloids. Tropane alkaloids are
634 characterized by an eight-membered, bicyclic, nitrogen-containing core and their synthesis is reported in
635 10 plant families, separated by ~120 Mya of evolution[130]. For example, the well-known narcotic cocaine is
636 synthesized by *Erythroxylum coca* (Erythroxylaceae) while cochlearine is synthesized in *Cochlearia*
637 *officinalis* (Brassicaceae). The Solanaceae family has emerged as a model system for studying tropane
638 alkaloid biosynthesis, but comparative studies reveal instances of independent evolution of tropanes in
639 distinct plant lineages[131, 132].
640
641 Scopolamine and hyoscyamine are tropane aromatic esters specific to the Solanaceae, and these
642 compounds derive their medicinal properties from anticholinergic effects, blocking activity of the
643 neurotransmitter acetylcholine. Scopolamine is used to treat a variety of illnesses including motion
644 sickness, drooling, and for palliative care in Parkinson's disease[133-135]. Tropane aromatic ester production
645 requires the biosynthesis of the tropane core as well as condensation of a phenyllactic acid moiety
646 through an ester linkage[136]. Although the biosynthesis of the tropane core intermediate and
647 polyhydroxylated derivates, known as calystegines, occurs in many genera of the Solanaceae, including
648 *Solanum*, the biosynthesis of tropane aromatic esters is restricted to the genera described above,
649 suggesting that not all species in the family possess the genes required for their synthesis[137]. Due to their
650 medicinal importance, considerable effort has focused on understanding the biosynthesis of hyoscyamine
651 and scopolamine.

652

653 Research leading to the elucidation of scopolamine biosynthesis spanned several decades, with progress
654 driven by the available technologies of the time. Initially, approaches focused on feeding labeled forms of
655 potential precursors to tropane producing plants and following incorporation of label into alkaloids[130].
656 This resulted in identification of pathway precursors and intermediates, as well as the development of an
657 overall framework of scopolamine biosynthesis. These efforts were followed by classical biochemical
658 approaches to purify enzymes based on activity. Peptide sequencing of the resulting purified enzymes
659 facilitated the design of oligonucleotide probes that were labeled and used to screen cDNA libraries to
660 identify the corresponding clones. Confirmation of function was achieved through characterization of
661 resulting recombinant enzymes expressed in *E. coli*. This led to the identification of several pathway genes,
662 including hyoscyamine 6β-hydroxylase *(H6H),* tropinone reductase I/II *(TRI* and *TRII*), and putrescine *N*-
663 methyltransferase *(PMT*). The development of expressed sequence tags in the mid-2000s, coupled with
664 virus-induced gene silencing (VIGS) for *in vivo* testing of function, led to the identification of littorine
665 mutase, an enzyme that catalyzes the rearrangement of littorine into hyoscyamine aldehyde[138]. More
666 recently, *Atropa belladonna* (Deadly Nightshade) emerged as a model for exploring tropane alkaloid
667 biosynthesis following the development of a multi-tissue transcriptome assembly and the deployment of
668 VIGS. These resources, coupled with synthetic biology, culminated in the identification of the missing
669 steps in scopolamine formation.

670

671 The first ring of the tropane core requires the conversion of ornithine, a non-proteinogenic amino acid,
672 into putrescine by ornithine decarboxylase (ODC). Putrescine is then *N*-methylated by putrescine
673 methyltransferase (PMT) and oxidized by methylputrescine oxidase (MPO). The *N*-methyl-Δ^1-pyrrolinium
674 cation forms through the spontaneous cyclization of *N*-methylaminobutanal, the product of MPO catalysis
675 (Figure 7). PMT requires *S*-adenosyl-L-methionine (SAM) to *N*-methylate putrescine and shares high
676 sequence similarity with spermidine synthase (*SPDS*), an enzyme involved in transferring the aminopropyl
677 moiety from decarboxylated SAM (dcSAM) onto putrescine to form spermidine, a ubiquitous
678 polyamine[139, 140]. It was hypothesized that *PMT* evolved from a gene duplication of *SPDS* and subsequent
679 neofunctionalization, and although SPDS cannot catalyze putrescine *N*-methylation, mutation of a single
680 SPDS amino acid, D103I, is sufficient to generate PMT activity[139]. The pyrrole moiety of nicotine, a natural
681 product produced in the *Nicotiana* genus of the Solanaceae, also requires *N*-methyl-Δ^1-pyrrolinium cation
682 biosynthesis. The biosynthetic steps leading to *N*-methyl-Δ^1-pyrrolinium cation formation are conserved
683 in *Nicotiana*, *Solanum*, and *Petunia* allowing the *N*-methyl-Δ^1-pyrrolinium cation to act as a core for

684 nicotine and tropane alkaloid biosynthesis found in Solanaceae and Convolvulaceae[141, 142]. In contrast, the
685 genes involved in the formation of the pyridine ring in nicotine biosynthesis are *Nicotiana*-specific
686 indicating that divergent evolution led to the formation of nicotine, likely through the duplication of the
687 genes in the nicotinamide adenine dinucleotide (NAD) cofactor biosynthetic pathway[142].
688
689 Formation of the tropane core in Solanaceae species requires a second cyclization event that yields
690 tropinone, which possesses a ketone functional group at the carbon-3 position of the core (Figure 7). The
691 first step in tropinone formation is catalyzed by a type III polyketide synthase, PYKS, which uses the *N*-
692 methyl-Δ^1-pyrrolinium cation and malonyl-Coenzyme A to form 4-(1-methyl-2-pyrrolidinyl)-3-oxobutanoic
693 acid[143]. Although PYKS can form 3-oxoglutaric acid without the *N*-methyl-Δ^1-pyrrolinium cation and these
694 two products can react non-enzymatically, the exact mechanism of 4-(1-methyl-2-pyrrolidinyl)-3-
695 oxobutanoic acid formation remains unclear[144, 145]. Tropinone synthase (CYP82M3) converts 4-(1-methyl-
696 2-pyrrolidinyl)-3-oxobutanoic acid to tropinone[143] Although putative orthologs of PYKS and CYP82M3 are
697 present in the genomes of several calystegine producing Solanaceous species including tomato, potato,
698 and pepper, these genes are absent in *Nicotiana* spp.; this is consistent with the lack of detectable
699 tropanes in these species[143]. In the Solanaceae, tropinone reductases I and II are members of the short-
700 chain dehydrogenase/reductase superfamily (SDR) that catalyze the reduction of the ketone of tropinone
701 to an alcohol to form tropine (3α-hydroxytropine) and pseudotropine (3β-hydroxytropine),
702 respectively[146]. TRI and TRII constitute a branch point in the tropane alkaloid biosynthetic pathway due to
703 their stereospecificity: TRI leads to the production of tropane aromatic esters, including hyoscyamine and
704 scopolamine and TRII directs flux towards calystegine production.
705
706 Biosynthesis of the principal aromatic tropane esters in the Solanaceae, littorine, hyoscyamine, and
707 scopolamine, requires the diversion of phenylalanine into the tropane pathway through a two-step
708 process that yields phenyllactic acid[147, 148](Figure 8). Identification of the aromatic aminotransferase
709 (AbArAT4) responsible for conversion of phenylalanine into phenylpyruvate revealed the power of
710 transcriptomics in Solanaceae tropane alkaloid enzyme discovery[147]. Analogous to bacterial aromatic
711 amino acid biosynthesis, a cytosolic aromatic aminotransferase from petunia (Ph-PPY-AT) catalyzes the
712 formation of phenylalanine from phenylpyruvate using tyrosine as an amino donor and yielding 4-
713 hydroxyphenylpyruvate[149]. AbArAT4 is related to Ph-PPY-AT and utilizes the same four substrates, but the
714 Atropa enzyme diverts phenylalanine into the tropane pathway by virtue of a ~250-fold more active
715 reverse reaction that yields phenylpyruvate and tyrosine. *AbArAT4* is co-expressed in the roots with other

716 tropane-related genes, and while silencing of this gene disrupts tropane alkaloid biosynthesis, it does not
717 alter aromatic amino acid pools, further supporting its neofunctionalized and specific role in specialized
718 metabolism[147]. Littorine biosynthesis requires the glycosylation of phenyllactate by a UDP-glucose
719 dependent glycosyltransferase followed by the acylation of tropine. The serine carboxypeptidase-like
720 (SCPL) acyltransferase (littorine synthase) acylates tropine using glycosylated phenyllactate as the acyl
721 donor[136].
722
723 Synthetic biology recently was utilized both to engineer scopolamine production in yeast and facilitate
724 the discovery of the final missing enzyme in the pathway, which had eluded discovery using *in planta*
725 experiments. The conversion of littorine to scopolamine requires four steps catalyzed by three enzymes
726 (Figure 8). Littorine mutase, a cytochrome P450, catalyzes the rearrangement of littorine to hyoscyamine
727 aldehyde[138], which is converted to hyoscyamine by hyoscyamine aldehyde dehydrogenase. Finally,
728 hyoscyamine-6-hydroxylase catalyzes the two-step hydroxylation and epoxidation of hyoscyamine to
729 scopolamine[150]. The production of scopolamine in yeast was achieved through the introduction of tropane
730 alkaloid pathway genes from several species, including *Datura stramonium*, *Datura metel*, and *Atropa*
731 *belladonna*[151]. Optimization of scopolamine production in yeast required the elimination of several native
732 genes to reduce the flow of tropane alkaloid intermediates into side products and the introduction of a
733 transporter from *Nicotiana tabacum* to facilitate transport of tropine into the vacuole for esterification
734 with phenyllactic acid[151]. Notably, the introduction of the pathway into yeast revealed the dehydrogenase
735 responsible for the reduction of hyoscyamine aldehyde into hyoscyamine, which had not previously been
736 identified *in planta*[151]. For example, silencing of this gene in *A. belladonna* did not result in a decrease in
737 downstream tropane alkaloids, likely due to promiscuous enzymatic activity of other dehydrogenases[152].
738 Hence, reconstruction of the pathway in a genetic host where background activities were removed
739 facilitated the identification of the final missing step in the scopolamine pathway.
740
741 **7.1. Independent evolution of tropanes in distinct plant lineages**
742 Evidence for independent evolution of tropanes in distinct plant lineages is manifest at different steps
743 throughout the pathway (Figures 7 & 8). While separate TRI and TRII enzymes reduce tropinone to tropine
744 or pseudotropine in the Solanaceae, a single SDR enzyme catalyzes both reactions in *C. officinalis*,
745 ultimately leading to tropine-derived cochlearine and pseudotropine-derived calystegines[131]. In addition,
746 while Solanaceae and Brassicaceae species utilize enzymes in the SDR family for the reduction of
747 tropinone, the analogous reaction in *E. coca* cocaine biosynthesis, the reduction of methylecgonone to

748 methylecgonine, is catalyzed by methylecgonone reductase (MecgoR) a member of the aldo-keto
749 reductase family [132]. Similarly, aromatic tropane ester biosynthesis is catalyzed by different classes of
750 acyltransferases in the Solanaceae and Erythroxylaceae. Littorine formation is synthesized by an SCPL
751 acyltransferase while cocaine synthase, which catalyzes the condensation of methylecgonine and
752 benzoyl-CoA, is a member of the BAHD acyltransferase family[153]. As additional tropane pathways in
753 distinct plant lineages are elucidated it is likely that further examples of independent evolution will be
754 discovered.
755
756 **8. Challenges and unexplored frontiers in Solanaceae metabolism.**
757 There has been a rapid increase in understanding the biosynthesis and evolution of plant SM pathways
758 during the last decade. Advances in genomics enabled gene-metabolite correlations in model and non-
759 model species. These data - combined with development of methods to test gene function in diverse
760 species, and transient expression in *Nicotiana benthamiana*, as well as engineering production in
761 microbial systems - led to the elucidation of multiple plant SM pathways and identified regulators of
762 known SM pathways [2-4, 151, 154]. The widespread adoption of these approaches, coupled with phylogeny-
763 guided comparative genomics and metabolomics, enabled exploration of the evolutionary trajectories of
764 the exemplary Solanaceae SM pathways described here.
765
766 However, despite advances in understanding Solanaceae SM biosynthesis and evolution, knowledge gaps
767 persist related to specific aspects of these well-studied pathways and opportunities exist to develop a
768 more comprehensive understanding of these pathways and networks. As evidenced through studies of
769 acylsugar evolution, much can be learned through adopting a broader sampling strategy to include more
770 phylogenetically diverse species that are typically less well studied[37, 72, 78]. Similar, phylogenetic-guided
771 metabolite screening approaches could be adopted to assess chemical diversity in other SM classes as the
772 foundation for exploring metabolite evolution using comparative genomics. For example, given the
773 tremendous chemical variation observed in trichome-derived acylsugars across the Solanaceae, and that
774 novel acylsugars were recently identified in root and root-exudates of tomato[48], it will be intriguing to
775 determine whether comparable root acylsugar diversity exists across the family and if so, to assess how
776 this diversity evolved.
777
778 There are also several examples where the biosynthesis of exemplary SM pathways in the Solanaceae are
779 not fully resolved. For example, the enzymes that catalyze the early steps in acylinositol biosynthesis in

780 *Solanum* spp. are yet to be reported. Similarly, the majority of the enzymes involved in capsaicinoid
781 biosynthesis and the final steps in nicotine biosynthesis await biochemical and functional
782 characterization[142, 155]. In addition, although the biosynthesis of scopolamine is elucidated and the
783 pathway reconstructed in yeast, the steps leading to the biosynthesis of other classes of Solanaceae
784 tropanes, including calystegines and schizanthines, are unknown[130, 156].
785
786 Comparative analyses of the evolution of SM-related gene clusters across the Solanaceae also remains
787 under-explored. For example, as outlined in this review, terpene and SGA-related gene clusters exist in
788 *Solanum* but variation across these clusters is mainly documented in a few model species, including
789 tomato, potato, eggplant, and closely related wild species[8, 82, 122]. Indeed, even for the comparatively well-
790 studied terpenoid-related gene clusters of tomato, many of the enzymes that reside within these clusters,
791 which may catalyze modifications of terpene scaffolds, remain uncharacterized. Furthermore, the extent
792 of conservation of terpene and other SM gene clusters across the Solanaceae is unknown. As multiple
793 chromosome scale genome assemblies of phylogenetically diverse Solanaceae species are available and
794 others will likely be generated soon, charting the evolutionary trajectories of SM gene clusters and the
795 metabolite variation they encode is now possible.
796
797 Finally, it is also worth noting that the most extensively characterized Solanaceae SM pathways are those
798 where the identities of the major metabolites were known for decades and their abundance is high in
799 specific cell types or tissues, facilitating purification and structural elucidation. It is more challenging to
800 identify unknown metabolites and purify metabolites that are of low abundance and technical challenges
801 persist that impede a more comprehensive understanding of metabolism and bridging of the gap between
802 genotype and phenotype.
803
804
805 **8.1. Challenges in the identification and annotation of SM enzymes.**
806 Advances in DNA sequencing are making development of chromosome-scale genome assemblies more
807 routine and recently several Solanaceae genomes were released, and the quality of existing assemblies
808 improved[17, 19, 157]. These studies allow the gene complement of an organism to be determined. However,
809 functional annotation of plant genomes remains incomplete, even for model species. The lack of accurate
810 annotation is particularly problematic for large gene families encoding SM-related enzymes that catalyze
811 common decorations of scaffold molecules, including cytochromes P450, 2-oxoglutarate dependent

812 dioxygenases, glycosyltransferases, and acyltransferases. SM-related enzymes are often catalytically
813 promiscuous and encoded by genes that evolved rapidly through duplication and associated
814 subfunctionalization, neofunctionalization, and gene loss[158]. Thus, annotation of SM enzymes based solely
815 on sequence similarity, predicted orthology, or synteny is often misleading. This concept is clearly
816 illustrated by examples identified through studying the evolution of acylsugar and terpene biosynthesis in
817 *Solanum* glandular trichomes. These studies reveal how activity can be altered by a few amino acid
818 differences in closely related enzymes from sister species, or diverse accessions within a species[64, 90, 95].
819 Hence, empirical determination of enzyme function remains imperative. Although characterization of
820 enzyme activities is often technically challenging, time consuming, and limited by substrate availability,
821 medium and high-throughput methods based on microtiter plates and microfluidics are utilized for
822 screening natural and computationally designed enzymes and such methods could potentially be adapted
823 for screening the activity of plant SM-related enzymes[159].
824
825 As documented throughout this review, co-expression is a powerful approach for predicting membership
826 of genes in metabolic pathways, particularly when there is *a priori* knowledge about enzymes from the
827 target pathway. Elucidation of the pathway leading to scopolamine biosynthesis, described above, is an
828 excellent example of the use of co-expression analyses to identify candidate genes co-expressed in roots.
829 However, when results of co-expression analysis are ambiguous or multiple candidate genes are
830 identified, as is often the case when investigating large SM-related gene families, additional filtering and
831 refinement of gene candidates may be required prior to time-consuming functional studies. In such cases,
832 comparative genomic analysis such as synteny or gene-cluster analysis - together with phylogenetic
833 analysis to determine whether gene candidates exhibit lineage-specific distribution or arose through a
834 recent duplication event - provide opportunities for refining candidate gene lists[160]. Outside of tomato,
835 there is a lack of publicly available transcriptome data, including data from diverse tissues, environmental
836 perturbations, and treatments. This limits novel metabolite pathway discovery in diverse Solanaceae
837 species and reduces the resolution of studies investigating the phylogenetic distribution and evolution of
838 SM pathways. Furthermore, plant SM pathways are often restricted to specific cell types, and therefore
839 the general focus on whole tissue sampling for transcriptome analysis can be limiting[68, 161, 162]. The recent
840 development of single-cell and single-nucleus transcriptome analyses holds great promise for increasing
841 the resolution of transcriptome data and refining candidate gene lists to facilitate the identification,
842 characterization, and cellular localization of Solanaceae SM pathways[163, 164].
843

844 Machine learning is another promising approach to distinguish GM and SM-related enzymes without prior
845 knowledge of pathway membership or gene-metabolite correlation information. Multiple features
846 including gene expression, transcriptional network analysis, rate of evolution, and duplication mechanism
847 allowed creation of statistical models that can distinguish GM from SM genes in *Arabidopsis*. In agreement
848 with the established characteristics of SM genes, machine learning models revealed that relative to GM
849 genes, SM genes tend to be less conserved, tandemly duplicated, more narrowly expressed, and
850 expressed at lower levels[165]. The prediction models also facilitated the classification of 1220 enzyme
851 encoding genes of unknown function as putatively SM-related. Similar machine learning strategies were
852 deployed in tomato to predict gene association with SM or GM pathways and to determine if gene
853 expression data can predict metabolic pathway membership[166, 167]. These approaches show potential to
854 build high-quality models but are limited by the quality of the input data, including mis-annotations and
855 the low number of functionally validated reference genes in tomato. These current limitations suggest
856 that application of machine learning for *de novo* prediction of novel SM pathways in tomato is not yet
857 possible at high accuracy. Furthermore, additional functional annotation, including the development of
858 more comprehensive genome and transcriptome data, will be needed to apply machine learning
859 approaches to predict SM pathway membership in additional members of the Solanaceae. Indeed, models
860 predicting whether a tomato gene is associated with specialized versus general metabolism were
861 improved when a transfer learning strategy was employed that utilized data from *Arabidopsis* models to
862 filter tomato annotations that disagreed with *Arabidopsis*[166] This represents a promising approach to using
863 comparative genomics data in specialized metabolic enzyme identification.
864

865 **8.2. Challenges in the identification and annotation of plant metabolites.**

866 Estimates suggest that ~10^6 metabolites are synthesized across species of the plant kingdom, collectively[1].
867 While we have deep knowledge of well-studied classes of plant metabolites, opportunities and challenges
868 for improving metabolome annotation remain. Several factors make separation and annotation of
869 metabolites challenging: for example, their diverse chemical composition, chemical properties (polarity
870 and hydrophilicity / hydrophobicity), and the orders of magnitude concentration range in which they
871 occur in biological samples[168, 169]. Improvements in analytical techniques, particularly liquid-
872 chromatography coupled with high-resolution mass-spectrometry (LC-HRMS) based metabolite profiling,
873 allows the detection of >10^3 metabolites within a single plant extract at high mass accuracy. However, a
874 single extraction solvent and chromatographic separation method are generally selected for individual
875 experiments, leading to unavoidable bias in the types of metabolites that are extracted and resolved and

876 therefore an under-representation of the metabolome[168]. Furthermore, most metabolites in a plant
877 extract are uncharacterized and many are of low abundance. In such cases, annotation can be challenging.
878 This is particularly true for specialized metabolites that are formed from diverse metabolic precursors,
879 possess multiple chemical modifications, and frequently exist as positional or structural isomers that may
880 be difficult to resolve. For example, even though tomato fruit ripening is one of the most extensively
881 studied plant biological processes, a large component of this metabolome remains unannotated. In a
882 recent study, untargeted metabolomics of tomato fruit at two different developmental stages identified
883 >1000 semi-lipophilic metabolites but only ~170 metabolites were annotated with some degree of
884 confidence, suggesting that the bulk of the tomato fruit metabolome remains unresolved[127]. Metabolite
885 databases containing spectra derived from tandem mass-spectrometry of known metabolites are
886 expanding and are useful for identifying unknown metabolites[170-172]. However, given the vast diversity of
887 plant metabolites and their frequent lineage-specific distribution, populating and curating such databases
888 requires substantial research funding, effort, and community engagement.
889
890 As with spatially resolved or single cell transcriptomics, the ability to obtain spatially resolved metabolome
891 data through mass spectrometry imaging of plant tissues represents an exciting development that will
892 enhance understanding of metabolism. Specifically, this technology will further refine the ability to detect
893 gene-metabolite correlations and allow the detection of metabolites that may be restricted to individual
894 cell types and therefore fall below the limit of detection in an extract prepared from a complex tissue
895 sample[173]. Mass spectrometry imaging has been utilized for investigating the spatial distribution of
896 metabolites in tomato fruit, including investigating the influence of genetic perturbation on SGA
897 accumulation[174]. Similarly, the spatial separation of SGAs and acylsugars were demonstrated in tomato
898 roots[48]. As improved MSI technologies develop and increase in availability, they will undoubtedly be more
899 widely adopted for exploring diverse aspects of Solanaceae metabolism.
900
901 Integration of genetic variation with metabolomics is a powerful approach to expand understanding of
902 SM metabolic networks and bridge the gap between genotype and phenotype. As described above, both
903 GWAS and metabolite QTL (mQTL) approaches were used to identify genomic regions and genes that
904 influence specialized metabolism in diverse tissues of tomato. In particular, the *S. lycopersicum* x *S.*
905 *pennellii* introgression line and the related backcross introgression line (BIL) populations were
906 foundational to improving understanding of the loci that influence metabolism within the tomato clade[33,
907 62, 117, 127, 175]. Approaches that harness natural variation are limited to species where it is possible to develop

908 inter-specific genetic populations or sufficient genetic variation is present within a species, to facilitate
909 GWAS. Although not currently as extensively characterized as the genetic resources for tomato,
910 germplasm panels and genetic populations, including introgression lines, are being developed and
911 characterized for the three additional major food crops of the Solanaceae; potato, pepper, and
912 eggplant[105, 176, 177]. In some cases, these genetic resources are being utilized to investigate metabolic
913 diversity via targeted and untargeted metabolomics and refinement of these efforts should facilitate
914 linking genotype to phenotype[178, 179].
915
916 An alternative, less frequently utilized, approach to harness genetic variation to interrogate metabolism
917 is to combine untargeted metabolite profiling with targeted disruption or over-expression of known
918 enzymes or transcription factors[180, 181]. This approach, while more targeted than a strategy incorporating
919 genome-wide genetic variation, can be utilized in any species where genetic manipulation is feasible and
920 has significant potential to increase understanding of plant SM networks. For example, disruption of an
921 SM enzyme will result in reduction of metabolites downstream of the enzyme, while the abundance of
922 metabolites upstream of the target enzyme can increase. This approach also allows detection of alternate
923 fates for pathway metabolites that accumulate due to gene disruption, revealing the existence of
924 biosynthetically linked metabolites. Referred to as "silent metabolism" this component of the
925 metabolome is likely substantial and certainly under-explored, including for engineering of novel
926 products[182]. Furthermore, as SM enzymes possess increased tendency for catalytic promiscuity,
927 untargeted metabolite profiling of lines disrupted in an enzyme of interest may reveal the existence of
928 previously uncharacterized catalytic activities.
929
930 While purification and structural elucidation of metabolites by NMR is a cornerstone of SM pathway
931 discovery, it is time-consuming and typically represents a major bottleneck. This is especially problematic
932 for metabolites that are of low abundance or co-purify with other compounds. Recent structural
933 elucidation of acyl-hexoses from *S. nigrum* was achieved using a combination of LC-MS, GC-MS, and 2D-
934 NMR approaches from crude and partially purified extracts without purification to homogeneity[72]. Similar
935 approaches should be adaptable to resolve the structures of other metabolites present in semi-purified
936 plant extracts. The recent adoption of microcrystal electron diffraction (MicroED) for structural
937 elucidation, including absolute stereochemistry, of mixtures of small organic molecules also shows great
938 promise for structural elucidation of plant specialized metabolites[183, 184]. MicroED can be used to resolve
939 the structures of nanocrystals of ~100 nm (~10^{-15} g) and thus is potentially more suitable for low

940 abundance metabolites than NMR, which typically requires hundreds of micrograms to milligram
941 quantities of purified compound. Application of this technology to specialized metabolite discovery was
942 recently demonstrated through a combined genome-mining, synthetic biology, and MicroED analysis that
943 elucidated the biosynthesis and structures of several 2-pyrridone metabolites from fungi[185]. Similarly,
944 synthetic biology can be utilized to engineer production of plant SMs in heterologous systems for
945 subsequent purification and structural elucidation. This strategy was effectively demonstrated by the
946 synthesis of gram scale quantities of the triterpene β-amyrin by vacuum infiltration of *N. benthamiana* co-
947 expressing a feedback insensitive variant of HMG-CoA reductase and oat β-amyrin synthase[186].
948 Subsequent experiments combining co-expression of these enzymes with triterpene decorating
949 cytochrome P450s from multiple species facilitated the production of novel non-natural triterpenes at
950 sufficient scale to allow purification and structural determination by NMR. *N. benthamiana* is widely used
951 for transient expression of candidate genes and as demonstrated above, represents a readily scalable
952 platform to produce metabolites for purification and subsequent structural elucidation.
953
954 **9. Conclusions**
955 Advances in genomics and metabolomics continue to enable greater understanding of SM pathway
956 biosynthesis and evolution. This review focused on the catalytic steps of five well-studied SM classes that
957 show varying degrees of lineage-specific distribution across the Solanaceae. This genetic variation,
958 coupled with high abundance, and often restricted distribution in specific tissue or cell types, facilitated
959 both purification and structural elucidation of these diverse metabolites as well as the identification of
960 the enzymes responsible for their biosynthesis. For example, acylsugar and terpene biosynthesis in
961 glandular trichomes, nicotine and tropane alkaloid biosynthesis in roots, and capsaicinoid biosynthesis in
962 pepper fruit placenta. These studies reveal examples of both intra- and inter-specific variation as well as
963 convergent evolution that has shaped the metabolic landscape across the Solanaceae. However, only a
964 small fraction of the metabolome and the genes responsible for its formation are resolved. Thus, many
965 opportunities exist to expand understanding of known pathways as well as identify novel pathways that
966 will enable a network level understanding of metabolism across the Solanaceae and identify target
967 molecules for agricultural and medicinal applications.
968
969 **10. Conflicts of interest**
970 There are no conflicts to declare.
971

11. Acknowledgements

We thank Dr. Yann-Ru Lou for providing the image of *Solanum nigrum*. Research on plant specialized metabolism in our labs is supported by National Science Foundation award numbers IOS-1546617, CBET-1565232 and MCB-1714093. C.S.B. and R.L.L. are supported in part by Michigan AgBioResearch and through USDA National Institute of Food and Agriculture, Hatch project numbers MICL02552 and MICL02458, respectively. This publication was made possible by Plant Biotechnology for Health and Sustainability predoctoral training awards to P.D.F. and H.M.P. from Grant Number T32-GM110523 from the National Institute of General Medical Sciences of the National Institutes of Health. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIGMS or NIH.

12. Notes and references

1. F. M. Afendi, T. Okada, M. Yamazaki, A. Hirai-Morita, Y. Nakamura, K. Nakamura, S. Ikeda, H. Takahashi, M. Altaf-Ul-Amin, L. K. Darusman, K. Saito and S. Kanaya, *Plant and Cell Physiology*, 2012, **53**.
2. R. S. Nett, W. Lau and E. S. Sattely, *Nature*, 2020, **584**, 148-153.
3. W. Lau and E. S. Sattely, *Science*, 2015, **349**, 1224-1228.
4. E. Fossati, A. Ekins, L. Narcross, Y. Zhu, J.-P. Falgueyret, G. A. W. Beaudoin, P. J. Facchini and V. J. J. Martin, *Nature Communications*, 2014, **5**, Article number 3283.
5. T. Vogt, *Molecular Plant*, 2010, **3**, 2-20.
6. C. A. Schenck and R. L. Last, *The FEBS Journal*, 2020, **287**, 1359-1368.
7. P. D. Sonawane, A. Jozwiak, S. Panda and A. Aharoni, *Current Opinion in Plant Biology*, 2020, **55**, 118-128.
8. M. Itkin, U. Heinig, O. Tzfadia, A. J. Bhide, B. Shinde, P. D. Cardenas, S. E. Bocobza, T. Unger, S. Malitsky, R. Finkers, Y. Tikunov, A. Bovy, Y. Chikate, P. Singh, I. Rogachev, J. Beekwilder, A. P. Giri and A. Aharoni, *Science*, 2013, **341**, 175-179.
9. B. J. Leong and R. L. Last, *Current Opinion in Structural Biology*, 2017, **47**, 105-112.
10. R. Akiyama, B. Watanabe, M. Nakayasu, H. J. Lee, J. Kato, N. Umemoto, T. Muranaka, K. Saito, Y. Sugimoto and M. Mizutani, *Nature Communications*, 2021, **12**, Article number 1300.
11. T. Särkinen, L. Bohs, R. G. Olmstead and S. Knapp, *BMC Evolutionary Biology*, 2013, **13**, Article Number 214.
12. E. Gagnon, R. Hilgenhof, A. Orejuela, A. McDonnell, G. Sablok, X. Aubriot, L. Giacomin, Y. Gouvêa, T. Bragionis, J. R. Stehmann, L. Bohs, S. Dodsworth, C. Martine, P. Poczai, S. Knapp and T. Särkinen, *American journal of botany*, 2022, DOI: 10.1002/ajb2.1827.
13. S. Knapp, L. Bohs, M. Nee and D. M. Spooner, *Comparative and Functional Genomics*, 2004, **5**, 285-291.
14. The Potato Genome Sequencing Consortium, *Nature*, 2011, **475**, 189-195.
15. The Tomato Genome Consortium, *Nature*, 2012, **485**, 635-641.
16. L. Gao, I. Gonda, H. Sun, Q. Ma, K. Bao, D. M. Tieman, E. A. Burzynski-Chang, T. L. Fish, K. A. Stromberg, G. L. Sacks, T. W. Thannhauser, M. R. Foolad, M. J. Diez, J. Blanca, J. Canizares, Y. Xu,

1012 E. van der Knaap, S. Huang, H. J. Klee, J. J. Giovannoni and Z. Fei, *Nature Genetics*, 2019, **51**,
1013 1044-1051.
1014 17. M. Alonge, X. Wang, M. Benoit, S. Soyk, L. Pereira, L. Zhang, H. Suresh, S. Ramakrishnan, F.
1015 Maumus, D. Ciren, Y. Levy, T. H. Harel, G. Shalev-Schlosser, Z. Amsellem, H. Razifard, A. L.
1016 Caicedo, D. M. Tieman, H. Klee, M. Kirsche, S. Aganezov, T. R. Ranallo-Benavidez, Z. H. Lemmon,
1017 J. Kim, G. Robitaille, M. Kramer, S. Goodwin, W. R. McCombie, S. Hutton, J. Van Eck, J. Gillis, Y.
1018 Eshed, F. J. Sedlazeck, E. van der Knaap, M. C. Schatz and Z. B. Lippman, *Cell*, 2020, **182**, 145-161.
1019 18. L. A. Mueller, T. H. Solow, N. Taylor, B. Skwarecki, R. Buels, J. Binns, C. Lin, M. H. Wright, R.
1020 Ahrens, Y. Wang, E. V. Herbst, E. R. Keyder, N. Menda, D. Zamir and S. D. Tanksley, *Plant*
1021 *Physiology*, 2005, **138**, 1310-1317.
1022 19. L. Barchi, M. T. Rabanus-Wallace, J. Prohens, L. Toppino, S. Padmarasu, E. Portis, G. L. Rotino, N.
1023 Stein, S. Lanteri and G. Giuliano, *The Plant Journal*, 2021, **107**, 579-596.
1024 20. B. Song, Y. Song, Y. Fu, E. B. Kizito, S. N. Kamenya, P. N. Kabod, H. Liu, S. Muthemba, R. Kariba, J.
1025 Njuguna, S. Maina, F. Stomeo, A. Djikeng, P. S. Hendre, X. Chen, W. Chen, X. Li, W. Sun, S. Wang,
1026 S. Cheng, A. Muchugi, R. Jamnadass, H.-Y. Shapiro, A. Van Deynze, H. Yang, J. Wang, X. Xu, D. A.
1027 Odeny and X. Liu, *Gigascience*, 2019, **8**, Article number giz115.
1028 21. Y. Eshed and D. Zamir, *Genetics*, 1995, **141**, 1147-1162.
1029 22. I. Ofner, J. Lashbrooke, T. Pleban, A. Aharoni and D. Zamir, *The Plant Journal*, 2016, **87**, 151-160.
1030 23. T. W. Toal, M. Ron, D. Gibson, K. Kajala, B. Splitt, L. S. Johnson, N. D. Miller, R. Slovak, A.
1031 Gaudinier, R. Patel, M. de Lucas, N. J. Provart, E. P. Spalding, W. Busch, D. J. Kliebenstein and S.
1032 M. Brady, *G3 Genes|Genomes|Genetics*, 2018, **8**, 3841-3855.
1033 24. A. L. Schilmiller, A. L. Charbonneau and R. L. Last, *Proceedings of the National Academy of*
1034 *Sciences of the United States of America*, 2012, **109**, 16377-16382.
1035 25. E. Fridman, F. Carrari, Y.-S. Liu, A. R. Fernie and D. Zamir, *Science*, 2004, **305**, 1786-1789.
1036 26. C. Brooks, V. Nekrasov, Z. B. Lippman and J. Van Eck, *Plant Physiology*, 2014, **166**, 1292-1297.
1037 27. Y. Liu, M. Schiff and S. P. Dinesh-Kumar, *The Plant Journal*, 2002, **31**, 777-786.
1038 28. E. G. W. M. Schijlen, C. H. R. de Vos, S. Martens, H. H. Jonker, F. M. Rosin, J. W. Molthoff, Y. M.
1039 Tikunov, G. C. Angenent, A. J. van Tunen and A. G. Bovy, *Plant Physiology*, 2007, **144**, 1520-1530.
1040 29. J. Van Eck, *Current Opinion in Biotechnology*, 2018, **49**, 35-41.
1041 30. D. Tieman, G. Zhu, M. F. R. Resende, Jr., T. Lin, M. Taylor, B. Zhang, H. Ikeda, Z. Liu, J. Fisher, I.
1042 Zemach, A. Monforte, D. Zamir, A. Granell, M. Kirst, S. Huang, H. Klee, C. Nguyen, D. Bies, J. L.
1043 Rambla and K. S. O. Beltran, *Science*, 2017, **355**, 391-394.
1044 31. G. Zhu, S. Wang, Z. Huang, S. Zhang, Q. Liao, C. Zhang, T. Lin, M. Qin, M. Peng, C. Yang, X. Cao, X.
1045 Han, X. Wang, E. van der Knaap, Z. Zhang, X. Cui, H. Klee, A. R. Fernie, J. Luo and S. Huang, *Cell*,
1046 2018, **172**, 249-261.
1047 32. X. Li, D. Tieman, Z. Liu, K. Chen and H. J. Klee, *The Plant Journal*, 2020, **104**, 631-644.
1048 33. K. Garbowicz, Z. Liu, S. Alseekh, D. Tieman, M. Taylor, A. Kuhalskaya, I. Ofner, D. Zamir, H. J. Klee,
1049 A. R. Fernie and Y. Brotman, *Molecular Plant*, 2018, **11**, 1147-1165.
1050 34. M. I. Zanor, J.-L. Rambla, J. Chaib, A. Steppa, A. Medina, A. Granell, A. R. Fernie and M. Causse,
1051 *Journal of Experimental Botany*, 2009, **60**, 2139-2154.
1052 35. Y. M. Tikunov, J. Molthoff, R. C. H. de Vos, J. Beekwilder, A. van Houwelingen, J. J. J. van der
1053 Hooft, M. Nijenhuis-de Vries, C. W. Labrie, W. Verkerke, H. van de Geest, M. V. Zamora, S. Presa,
1054 J. Luis Rambla, A. Granell, R. D. Hall and A. G. Bovy, *Plant Cell*, 2013, **25**, 3067-3078.
1055 36. E. Maldonado, F. R. Torres, M. Martínez and A. L. Pérez-Castorena, *Journal of Natural Products*,
1056 2006, **69**, 1511-1513.
1057 37. G. D. Moghe, B. J. Leong, S. M. Hurney, A. D. Jones and R. L. Last, *eLife*, 2017, **6**, e28468.
1058 38. T. Asai and Y. Fujimoto, *Phytochemistry*, 2010, **71**, 1410-1417.
1059 39. T. Asai, N. Hara and Y. Fujimoto, *Phytochemistry*, 2010, **71**, 877-894.

1060 40. T. Asai and Y. Fujimoto, *Phytochemistry Letters*, 2011, **4**, 38-42.
1061 41. T. Asai, T. Sakai, K. Ohyama and Y. Fujimoto, *Chemical & Pharmaceutical Bulletin*, 2011, **59**, 747-
1062 752.
1063 42. M. Bah and R. Pereda-Miranda, *Tetrahedron*, 1996, **52**, 13063-13080.
1064 43. R. Pereda-Miranda, R. Mata, A. L. Anaya, D. B. M. Wickramaratne, J. M. Pezzuto and A. D.
1065 Kinghorn, *Journal of Natural Products*, 1993, **56**, 571-582.
1066 44. M. Ono, A. Takigawa, H. Muto, K. Kabata, M. Okawa, J. Kinjo, K. Yokomizo, H. Yoshimitsu and T.
1067 Nohara, *Chemical & Pharmaceutical Bulletin*, 2015, **63**, 641-648.
1068 45. Q. Wu, J.-G. Cho, D.-S. Lee, D.-Y. Lee, N.-Y. Song, Y.-C. Kim, K.-T. Lee, H.-G. Chung, M.-S. Choi, T.-
1069 S. Jeong, E.-M. Ahn, G.-S. Kim and N.-I. Baek, *Carbohydrate Research*, 2013, **372**, 9-14.
1070 46. C. Li, Z. Wang and A. D. Jones, *Analytical and Bioanalytical Chemistry*, 2014, **406**, 171-182.
1071 47. T. Nakashima, H. Wada, S. Morita, R. Erra-Balsells, K. Hiraoka and H. Nonami, *Analytical*
1072 *Chemistry*, 2016, **88**, 3049-3057.
1073 48. E. Korenblum, Y. Dong, J. Szymanski, S. Panda, A. Jozwiak, H. Massalha, S. Meir, I. Rogachev and
1074 A. Aharoni, *Proceedings of the National Academy of Sciences of the United States of America*,
1075 2020, **117**, 3874-3883.
1076 49. B. Ghosh, T. C. Westbrook and A. D. Jones, *Metabolomics*, 2014, **10**, 496-507.
1077 50. X. Liu, M. Enright, C. S. Barry and A. D. Jones, *Metabolomics*, 2017, **13**.
1078 51. C.-R. Zhang, W. Khan, J. Bakht and M. G. Nair, *Food Chemistry*, 2016, **196**, 726-732.
1079 52. C.-Y. Zhang, J.-G. Luo, R.-H. Liu, R. Lin, M.-H. Yang and L.-Y. Kong, *Fitoterapia*, 2016, **114**, 138-
1080 143.
1081 53. D. B. Lybrand, T. M. Anthony, A. D. Jones and R. L. Last, *Metabolites*, 2020, **10**, Article number
1082 401.
1083 54. E. Cicchetti, L. Duroure, E. Le Borgne and R. Laville, *Journal of Natural Products*, 2018, **81**, 1946-
1084 1955.
1085 55. C.-M. Cao, X. Wu, K. Kindscher, L. Xu and B. N. Timmermann, *Journal of Natural Products*, 2015,
1086 **78**, 2488-2493.
1087 56. C.-A. Bernal, L. Castellanos, D. M. Aragón, D. Martínez-Matamoros, C. Jiménez, Y. Baena and F.
1088 A. Ramos, *Carbohydrate Research*, 2018, **461**, 4-10.
1089 57. S. M. Hurney, PhD, Michigan State University, 2018.
1090 58. B. M. Leckie, D. A. D'Ambrosio, T. M. Chappell, R. Halitschke, D. M. De Jong, A. Kessler, G. G.
1091 Kennedy and M. A. Mutschler, *PLOS ONE*, 2016, **11**, Article number e0153345.
1092 59. J. C. Goffreda, M. A. Mutschler, D. A. Avé, W. M. Tingey and J. C. Steffens, *Journal of Chemical*
1093 *Ecology*, 1989, **15**, 2135-2147.
1094 60. L. Van Thi, A. Weinhold, C. Ullah, S. Dressel, M. Schoettner, K. Gase, E. Gaquerel, S. Xu and I. T.
1095 Baldwin, *Plant Physiology*, 2017, **174**, 370-386.
1096 61. A. Weinhold and I. T. Baldwin, *Proceedings of the National Academy of Sciences of the United*
1097 *States of America*, 2011, **108**, 7855-7859.
1098 62. A. Schilmiller, F. Shi, J. Kim, A. L. Charbonneau, D. Holmes, A. D. Jones and R. L. Last, *The Plant*
1099 *Journal*, 2010, **62**, 391-403.
1100 63. A. L. Schilmiller, G. D. Moghe, P. Fan, B. Ghosh, J. Ning, A. D. Jones and R. L. Last, *Plant Cell*,
1101 2015, **27**, 1002-1017.
1102 64. P. Fan, A. M. Miller, A. L. Schilmiller, X. Liu, I. Ofner, A. D. Jones, D. Zamir and R. L. Last,
1103 *Proceedings of the National Academy of Sciences of the United States of America*, 2016, **113**,
1104 E239-E248.
1105 65. J. Kim, K. Kang, E. Gonzales-Vigil, F. Shi, A. D. Jones, C. S. Barry and R. L. Last, *Plant Physiology*,
1106 2012, **160**, 1854-1870.

1107 66. J. B. Landis, C. M. Miller, A. K. Broz, A. A. Bennett, N. Carrasquilla-Garcia, D. R. Cook, R. L. Last, P.
1108 A. Bedinger and G. D. Moghe, *Molecular Biology and Evolution*, 2021, **38**, 3202-3219.
1109 67. P. Fan, A. M. Miller, X. Liu, A. D. Jones and R. L. Last, *Nature Communications*, 2017, **8**, Article
1110 number 2080.
1111 68. B. J. Leong, D. B. Lybrand, Y.-R. Lou, P. Fan, A. L. Schilmiller and R. L. Last, *Science Advances*,
1112 2019, **5**, eaaw3754.
1113 69. J. Ning, G. D. Moghe, B. Leong, J. Kim, I. Ofner, Z. Wang, C. Adams, A. D. Jones, D. Zamir and R. L.
1114 Last, *Plant Physiology*, 2015, **169**, 1821-1835.
1115 70. P. Fan, P. Wang, Y.-R. Lou, B. J. Leong, B. M. Moore, C. A. Schenck, R. Combs, P. Cao, F. Brandizzi,
1116 S.-H. Shiu and R. L. Last, *eLife*, 2020, **9**, Article number e56717.
1117 71. B. J. Leong, S. M. Hurney, P. D. Fiesel, G. D. Moghe, A. D. Jones and R. L. Last, *Plant Physiology*,
1118 2020, **183**, 915-924.
1119 72. Y.-R. Lou, T. M. Anthony, P. D. Fiesel, R. E. Arking, E. M. Christensen, A. D. Jones and R. L. Last,
1120 *Science Advances*, 2021, **7**, eabj8726.
1121 73. O. T. Chortyk, S. J. Kays and Q. Teng, *Journal of Agricultural and Food Chemistry*, 1997, **45**, 270-
1122 275.
1123 74. O. T. Chortyk, R. F. Severson, H. C. Cutler and V. A. Sisson, *Bioscience Biotechnology and*
1124 *Biochemistry*, 1993, **57**, 1355-1356.
1125 75. R. R. King and L. A. Calhoun, *Phytochemistry*, 1988, **27**, 3761-3763.
1126 76. T. Matsuzaki, Y. Shinozaki, S. Suhara, M. Ninomiya, H. Shigematsu and A. Koiwai, *Agricultural*
1127 *and Biological Chemistry*, 1989, **53**, 3079-3082.
1128 77. Y. Herrera-Salgado, M. L. Garduño-Ramírez, L. Vázquez, M. Y. Rios and L. Alvarez, *Journal of*
1129 *Natural Products*, 2005, **68**, 1031-1036.
1130 78. S. S. Nadakuduti, J. B. Uebler, X. Liu, A. D. Jones and C. S. Barry, *Plant Physiology*, 2017, **175**, 36-
1131 50.
1132 79. F. Zhou and E. Pichersky, *Current Opinion in Plant Biology*, 2020, **55**, 1-10.
1133 80. L. Pazouki and U. Niinemets, *Frontiers in Plant Science*, 2016, **7**, Article number 1019.
1134 81. P. S. Karunanithi and P. Zerbe, *Frontiers in Plant Science*, 2019, **10**, Article number 1166.
1135 82. F. Zhou and E. Pichersky, *New Phytologist*, 2020, **226**, 1341-1360.
1136 83. T. A. Akhtar, Y. Matsuba, I. Schauvinhold, G. Yu, H. A. Lees, S. E. Klein and E. Pichersky, *The Plant*
1137 *Journal*, 2013, **73**, 640-652.
1138 84. A. M. Boutanaev, T. Moses, J. Zi, D. R. Nelson, S. T. Mugford, R. J. Peters and A. Osbourn,
1139 *Proceedings of the National Academy of Sciences of the United States of America*, 2015, **112**,
1140 E81-E88.
1141 85. Y. Matsuba, T. T. H. Nguyen, K. Wiegert, V. Falara, E. Gonzales-Vigil, B. Leong, P. Schaefer, D.
1142 Kudrna, R. A. Wing, A. M. Bolger, B. Usadel, A. Tissier, A. R. Fernie, C. S. Barry and E. Pichersky,
1143 *Plant Cell*, 2013, **25**, 2022-2036.
1144 86. J. Zi, Y. Matsuba, Y. J. Hong, A. J. Jackson, D. J. Tantillo, E. Pichersky and R. J. Peters, *Journal of*
1145 *the American Chemical Society*, 2014, **136**, 16951-16953.
1146 87. E. Gonzales-Vigil, D. E. Hufnagel, J. Kim, R. L. Last and C. S. Barry, *The Plant Journal*, 2012, **71**,
1147 921-935.
1148 88. A. L. Schilmiller, I. Schauvinhold, M. Larson, R. Xu, A. L. Charbonneau, A. Schmidt, C. Wilkerson,
1149 R. L. Last and E. Pichersky, *Proceedings of the National Academy of Sciences of the United States*
1150 *of America*, 2009, **106**, 10865-10870.
1151 89. C. Sallaud, D. Rontein, S. Onillon, F. Jabès, P. Duffé, C. Giacalone, S. Thoraval, C. Escoffier, G.
1152 Herbette, N. Leonhardt, M. Causse and A. Tissier, *Plant Cell*, 2009, **21**, 301-317.
1153 90. J.-H. Kang, E. Gonzales-Vigil, Y. Matsuba, E. Pichersky and C. S. Barry, *Plant Physiology*, 2014,
1154 **164**, 80-91.

1155 91. P. M. Bleeker, P. J. Diergaarde, K. Ament, S. Schütz, B. Johne, J. Dijkink, H. Hiemstra, R. de
1156 Gelder, M. T. J. de Both, M. W. Sabelis, M. A. Haring and R. C. Schuurink, *Phytochemistry*, 2011,
1157 **72**, 68-73.
1158 92. P. M. Bleeker, R. Mirabella, P. J. Diergaarde, A. VanDoorn, A. Tissier, M. R. Kant, M. Prins, M. de
1159 Vos, M. A. Haring and R. C. Schuurink, *Proceedings of the National Academy of Sciences of the*
1160 *United States of America*, 2012, **109**, 20124-20129.
1161 93. R. M. Coates, J. F. Denissen, J. A. Juvik and B. A. Babka, *Journal of Organic Chemistry*, 1988, **53**,
1162 2186-2192.
1163 94. J. E. Frelichowski and J. A. Juvik, *Journal of Economic Entomology*, 2001, **94**, 1249-1259.
1164 95. S. Zabel, W. Brandt, A. Porzel, B. Athmer, S. Bennewitz, P. Schäfer, R. Kortbeek, P. Bleeker and A.
1165 Tissier, *The Plant Journal*, 2021, **105**, 1309-1325.
1166 96. H. Duranova, V. Valkova and L. Gabriny, *Phytochemistry Reviews*, DOI: 10.1007/s11101-021-
1167 09789-7.
1168 97. J. R. Friedman, S. D. Richbart, J. C. Merritt, K. C. Brown, K. L. Denning, M. T. Tirona, M. A.
1169 Valentovic, S. L. Miles and P. Dasgupta, *Biomedicine & Pharmacotherapy*, 2019, **118**.
1170 98. F. Spiller, M. K. Alves, S. M. Vieira, T. A. Carvalho, C. E. Leite, A. Lunardelli, J. A. Poloni, F. Q.
1171 Cunha and J. R. de Oliveira, *Journal of Pharmacy and Pharmacology*, 2008, **60**, 473-478.
1172 99. S. Varghese, P. Kubatka, L. Rodrigo, K. Gazdikova, M. Caprnda, J. Fedotova, A. Zulli, P. Kruzliak
1173 and D. Büsselberg, *International Journal of Food Sciences and Nutrition*, 2017, **68**, 392-401.
1174 100. J. J. Tewksbury and G. P. Nabhan, *Nature*, 2001, **412**, 403-404.
1175 101. M. J. Caterina, A. Leffler, A. B. Malmberg, W. J. Martin, J. Trafton, K. R. Petersen-Zeitz, M.
1176 Koltzenburg, A. I. Basbaum and D. Julius, *Science*, 2000, **288**, 306-313.
1177 102. S. E. Jordt and D. Julius, *Cell*, 2002, **108**, 421-430.
1178 103. C. C. García, M. H. J. Barfuss, E. M. Sehr, G. E. Barboza, R. Samuel, E. A. Moscone and F.
1179 Ehrendorfer, *Annals of Botany*, 2016, **118**, 35-51.
1180 104. S. Kim, M. Park, S.-I. Yeom, Y.-M. Kim, J. M. Lee, H.-A. Lee, E. Seo, J. Choi, K. Cheong, K.-T. Kim, K.
1181 Jung, G.-W. Lee, S.-K. Oh, C. Bae, S.-B. Kim, H.-Y. Lee, S.-Y. Kim, M.-S. Kim, B.-C. Kang, Y. D. Jo, H.-
1182 B. Yang, H.-J. Jeong, W.-H. Kang, J.-K. Kwon, C. Shin, J. Y. Lim, J. H. Park, J. H. Huh, J.-S. Kim, B.-D.
1183 Kim, O. Cohen, I. Paran, M. C. Suh, S. B. Lee, Y.-K. Kim, Y. Shin, S.-J. Noh, J. Park, Y. S. Seo, S.-Y.
1184 Kwon, H. A. Kim, J. M. Park, H.-J. Kim, S.-B. Choi, P. W. Bosland, G. Reeves, S.-H. Jo, B.-W. Lee, H.-
1185 T. Cho, H.-S. Choi, M.-S. Lee, Y. Yu, Y. Do Choi, B.-S. Park, A. van Deynze, H. Ashrafi, T. Hill, W. T.
1186 Kim, H.-S. Pai, H. K. Ahn, I. Yeam, J. J. Giovannoni, J. K. C. Rose, I. Sørensen, S.-J. Lee, R. W. Kim,
1187 I.-Y. Choi, B.-S. Choi, J.-S. Lim, Y.-H. Lee and D. Choi, *Nature Genetics*, 2014, **46**, 270-278.
1188 105. P. Tripodi, M. T. Rabanus-Wallace, L. Barchi, S. Kale, S. Esposito, A. Acquadro, R. Schafleitner, M.
1189 van Zonneveld, J. Prohens, M. J. Diez, A. Börner, J. Salinier, B. Caromel, A. Bovy, F. Boyaci, G.
1190 Pasev, R. Brandt, A. Himmelbach, E. Portis, R. Finkers, S. Lanteri, I. Paran, V. Lefebvre, G.
1191 Giuliano and N. Stein, *Proceedings of the National Academy of Sciences of the United States of*
1192 *America*, 2021, **118**, Article number e2104315118.
1193 106. K. Han, H. Y. Lee, N. Y. Ro, O. S. Hur, J. H. Lee, J. K. Kwon and B. C. Kang, *Plant Biotechnology*
1194 *Journal*, 2018, **16**, 1546-1558.
1195 107. M. Park, J. H. Lee, K. Han, S. Jang, J. Han, J. H. Lim, J. W. Jung and B. C. Kang, *Theoretical and*
1196 *Applied Genetics*, 2019, **132**, 515-529.
1197 108. Y. Q. Lang, H. Kisaka, R. Sugiyama, K. Nomura, A. Morita, T. Watanabe, Y. Tanaka, S. Yazawa and
1198 T. Miwa, *The Plant Journal*, 2009, **59**, 953-961.
1199 109. Y. Tanaka, T. Sonoyama, Y. Muraga, S. Koeda, T. Goto, Y. Yoshida and K. Yasuba, *Molecular*
1200 *Breeding*, 2015, **35**.
1201 110. N. Weber, A. Ismail, M. Gorwa-Grauslund and M. Carlquist, *BMC Biotechnology*, 2014, **14**.

1202 111. S. Koeda, K. Sato, H. Saito, A. J. Nagano, M. Yasugi, H. Kudoh and Y. Tanaka, *Theoretical and*
1203 *Applied Genetics*, 2019, **132**, 65-80.
1204 112. C. Stewart, B. C. Kang, K. Liu, M. Mazourek, S. L. Moore, E. Y. Yoo, B. D. Kim, I. Paran and M. M.
1205 Jahn, *The Plant Journal*, 2005, **42**, 675-688.
1206 113. J. G. Roddick, M. Weissenberg and A. L. Leonard, *Phytochemistry*, 2001, **56**, 603-610.
1207 114. L. C. Dolan, R. A. Matulka and G. A. Burdock, *Toxins*, 2010, **2**, 2289-2332.
1208 115. X.-Y. Gu, X.-F. Shen, L. Wang, Z.-W. Wu, F. Li, B. Chen, G.-L. Zhang and M.-K. Wang,
1209 *Phytochemistry*, 2018, **147**, 125-131.
1210 116. Y. Iijima, B. Watanabe, R. Sasaki, M. Takenaka, H. Ono, N. Sakurai, N. Umemoto, H. Suzuki, D.
1211 Shibata and K. Aoki, *Phytochemistry*, 2013, **95**, 145-157.
1212 117. P. D. Cárdenas, P. D. Sonawane, U. Heinig, A. Jozwiak, S. Panda, B. Abebie, Y. Kazachkova, M.
1213 Pliner, T. Unger, D. Wolf, I. Ofner, E. Vilaprinyo, S. Meir, O. Davydov, A. Gal-On, S. Burdman, A.
1214 Giri, D. Zamir, T. Scherf, J. Szymanski, I. Rogachev and A. Aharoni, *Nature Communications*, 2019,
1215 **10**, Article number 5169.
1216 118. J. R. Paudel, C. Davidson, J. Song, I. Maxim, A. Aharoni and H. H. Tai, *Molecular Plant-Microbe*
1217 *Interactions*, 2017, **30**, 876-885.
1218 119. B. A. Shinde, B. B. Dholakia, K. Hussain, S. Panda, S. Meir, I. Rogachev, A. Aharoni, A. P. Giri and
1219 A. C. Kamble, *Plant Molecular Biology*, 2017, **95**, 411-423.
1220 120. M.-C. Sánchez-Mata, W. E. Yokoyama, Y.-J. Hong and J. Prohens, *Journal of Agricultural and Food*
1221 *Chemistry*, 2010, **58**, 5502-5508.
1222 121. P. D. Sonawane, J. Pollier, S. Panda, J. Szymanski, H. Massalha, M. Yona, T. Unger, S. Malitsky, P.
1223 Arendt, L. Pauwels, E. Almekias-Siegl, I. Rogachev, S. Meir, P. D. Cárdenas, A. Masri, M. Petrikov,
1224 H. Schaller, A. A. Schaffer, A. Kamble, A. P. Giri, A. Goossens and A. Aharoni, *Nature Plants*, 2017,
1225 **3**, Article 16205.
1226 122. L. Barchi, M. Pietrella, L. Venturini, A. Minio, L. Toppino, A. Acquadro, G. Andolfo, G. Aprea, C.
1227 Avanzato, L. Bassolino, C. Comino, A. Dal Molin, A. Ferrarini, L. C. Maor, E. Portis, S. Reyes-Chin-
1228 Wo, R. Rinaldi, T. Sala, D. Scaglione, P. Sonawane, P. Tononi, E. Almekias-Siegl, E. Zago, M. R.
1229 Ercolano, A. Aharoni, M. Delledonne, G. Giuliano, S. Lanteri and G. L. Rotino, *Scientific Reports*,
1230 2019, **9**, Article number 11769.
1231 123. P. D. Sonawane, U. Heinig, S. Panda, N. S. Gilboa, M. Yona, S. P. Kumar, N. Alkan, T. Unger, S.
1232 Bocobza, M. Pliner, S. Malitsky, M. Tkachev, S. Meir, I. Rogachev and A. Aharoni, *Proceedings of*
1233 *the National Academy of Sciences of the United States of America*, 2018, **115**, E5419-E5428.
1234 124. R. Akiyama, H. J. Lee, M. Nakayasu, K. Osakabe, Y. Osakabe, N. Umemoto, K. Saito, T. Muranaka,
1235 Y. Sugimoto and M. Mizutani, *Plant Biotechnology*, 2019, **36**, 253-263.
1236 125. H. J. Lee, M. Nakayasu, R. Akiyama, M. Kobayashi, H. Miyachi, Y. Sugimoto, N. Umemoto, K.
1237 Saito, T. Muranaka and M. Mizutani, *Plant and Cell Physiology*, 2019, **60**, 1304-1315.
1238 126. M. Nakayasu, R. Akiyama, M. Kobayashi, H. J. Lee, T. Kawasaki, B. Watanabe, S. Urakawa, J. Kato,
1239 Y. Sugimoto, Y. Iijima, K. Saito, T. Muranaka, N. Umemoto and M. Mizutani, *Plant and Cell*
1240 *Physiology*, 2020, **61**, 21-28.
1241 127. J. Szymański, S. Bocobza, S. Panda, P. Sonawane, P. D. Cárdenas, J. Lashbrooke, A. Kamble, N.
1242 Shahaf, S. Meir, A. Bovy, J. Beekwilder, Y. Tikunov, I. Romero de la Fuente, D. Zamir, I. Rogachev
1243 and A. Aharoni, *Nature Genetics*, 2020, **52**, 1111-1121.
1244 128. R. Akiyama, M. Nakayasu, N. Umemoto, J. Kato, M. Kobayashi, H. J. Lee, Y. Sugimoto, Y. Iijima, K.
1245 Saito, T. Muranaka and M. Mizutani, *Plant and Cell Physiology*, 2021, **62**, 775-783.
1246 129. Y. Kazachkova, I. Zemach, S. Panda, S. Bocobza, A. Vainer, I. Rogachev, Y. Dong, S. Ben-Dor, D.
1247 Veres, C. Kanstrup, S. K. Lambertz, C. Crocoll, Y. Hu, E. Shani, S. Michaeli, H. H. Nour-Eldin, D.
1248 Zamir and A. Aharoni, *Nature Plants*, 2021, **7**, 468 - 480.

130. N. Kim, O. Estrada, B. Chavez, Jr., C. Stewart and J. C. D'Auria, *Molecules*, 2016, **21**, Article number 1510.
131. A. Brock, W. Brandt and B. Dräeger, *The Plant Journal*, 2008, **54**, 388-401.
132. J. Jirschitzka, G. W. Schmidt, M. Reichelt, B. Schneider, J. Gershenzon and J. C. D'Auria, *Proceedings of the National Academy of Sciences of the United States of America*, 2012, **109**, 10304-10309.
133. A. Mato, J. Limeres, I. Tomás, M. Muñoz, C. Abuin, J. F. Feijoo and P. Diz, *British Journal of Clinical Pharmacology*, 2010, **69**, 684-688.
134. S. P. Clissold and R. C. Heel, *Drugs*, 1985, **29**, 189-207.
135. L. M. Pérez, C. Farriols, V. Puente, J. Planas and I. Ruiz, *Palliative Medicine*, 2011, **25**, 92-93.
136. F. Qiu, J. Zeng, J. Wang, J.-P. Huang, W. Zhou, C. Yang, X. Lan, M. Chen, S.-X. Huang, G. Kai and Z. Liao, *New Phytologist*, 2020, **225**, 1906-1914.
137. R. J. Nash, M. Rothschild, E. A. Porter, A. A. Watson, R. D. Waigh and P. G. Waterman, *Phytochemistry*, 1993, **34**, 1281-1283.
138. R. Li, D. W. Reed, E. Liu, J. Nowak, L. E. Pelcher, J. E. Page and P. S. Covello, *Chemistry & Biology*, 2006, **13**, 513-520.
139. A. Junker, J. Fischer, Y. Sichhart, W. Brandt and B. Dräger, *Frontiers in Plant Science*, 2013, **4**, Article number 260.
140. O. Stenzel, M. Teuber and B. Dräger, *Planta*, 2006, **223**, 200-212.
141. M. Kajikawa, N. Sierro, H. Kawaguchi, N. Bakaher, N. V. Ivanov, T. Hashimoto and T. Shoji, *Plant Physiology*, 2017, **174**, 999-1011.
142. S. Xu, T. Brockmöeller, A. Navarro-Quezada, H. Kuhl, K. Gase, Z. Ling, W. Zhou, C. Kreitzer, M. Stanke, H. Tang, E. Lyons, P. Pandey, S. P. Pandey, B. Timmermann, E. Gaquerel and I. T. Baldwin, *Proceedings of the National Academy of Sciences of the United States of America*, 2017, **114**, 6133-6138.
143. M. A. Bedewitz, A. D. Jones, J. C. D'Auria and C. S. Barry, *Nature Communications*, 2018, **9**, Article number 5281.
144. J.-P. Huang, C. Fang, X. Ma, L. Wang, J. Yang, J. Luo, Y. Yan, Y. Zhang and S.-X. Huang, *Nature Communications*, 2019, **10**, Article number 4036.
145. R. S. Nett, Y. Dho, Y.-Y. Low and E. S. Sattely, *Proceedings of the National Academy of Sciences of the United States of America*, 2021, **118**, Article number e2102949118.
146. K. Nakajima, T. Hashimoto and Y. Yamada, *Proceedings of the National Academy of Sciences of the United States of America*, 1993, **90**, 9591-9595.
147. M. A. Bedewitz, E. Góngora-Castillo, J. B. Uebler, E. Gonzales-Vigil, K. E. Wiegert-Rininger, K. L. Childs, J. P. Hamilton, B. Vaillancourt, Y.-S. Yeo, J. Chappell, D. DellaPenna, A. D. Jones, C. R. Buell and C. S. Barry, *Plant Cell*, 2014, **26**, 3745-3762.
148. F. Qiu, C. Yang, L. Yuan, D. Xiang, X. Lan, M. Chen and Z. Liao, *Organic Letters*, 2018, **20**, 7807-7810.
149. H. Yoo, J. R. Widhalm, Y. Qian, H. Maeda, B. R. Cooper, A. S. Jannasch, I. Gonda, E. Lewinsohn, D. Rhodes and N. Dudareva, *Nature Communications*, 2013, **4**, Article number 2833.
150. T. Hashimoto and Y. Yamada, *Plant Physiology*, 1986, **81**, 619-625.
151. P. Srinivasan and C. D. Smolke, *Nature*, 2020, **585**, 614-619.
152. F. Qiu, Y. Yan, J. Zeng, J.-P. Huang, L. Zeng, W. Zhong, X. Lan, M. Chen, S.-X. Huang and Z. Liao, *ACS Catalysis*, 2021, **11**, 2912-2924.
153. G. W. Schmidt, J. Jirschitzka, T. Porta, M. Reichelt, K. Luck, J. C. P. Torre, F. Dolke, E. Varesio, G. Hopfgartner, J. Gershenzon and J. C. D'Auria, *Plant Physiology*, 2015, **167**, 89-101.
154. Y. Li, Y. Chen, L. Zhou, S. You, H. Deng, Y. Chen, S. Alseekh, Y. Yuan, R. Fu, Z. Zhang, D. Su, A. R. Fernie, M. Bouzayen, T. Ma, M. Liu and Y. Zhang, *Molecular Plant*, 2020, **13**, 1203-1218.

1297 155. E. R. Naves, L. D. Silva, R. Sulpice, W. L. Araújo, A. Nunes-Nesi, L. E. P. Peres and A. Zsögön,
1298 *Trends in Plant Science*, 2019, **24**, 109-120.
1299 156. P. Christen, S. Cretton, M. Humam, S. Bieri, O. Muñoz and P. Joseph-Nathan, *Phytochemistry*
1300 *Reviews*, 2020, **19**, 615-641.
1301 157. T. P. Michael and R. VanBuren, *Current Opinion in Plant Biology*, 2020, **54**, 26-33.
1302 158. J.-K. Weng, R. N. Philippe and J. P. Noel, *Science*, 2012, **336**, 1667-1670.
1303 159. H. A. Bunzel, X. Garraboul, M. Pott and D. Hilvert, *Current Opinion in Structural Biology*, 2018,
1304 **48**, 149-156.
1305 160. J. R. Jacobowitz and J.-K. Weng, in *Annual Review of Plant Biology, Vol 71, 2020*, ed. S. S.
1306 Merchant, 2020, vol. 71, pp. 631-658.
1307 161. A. Onoyovwe, J. M. Hagel, X. Chen, M. F. Khan, D. C. Schriemer and P. J. Facchini, *Plant Cell*,
1308 2013, **25**, 4110-4122.
1309 162. V. Courdavault, N. Papon, M. Clastre, N. Giglioli-Guivarc'h, B. St-Pierre and V. Burlat, *Current*
1310 *Opinion in Plant Biology*, 2014, **19**, 43-50.
1311 163. K. H. Ryu, L. Huang, H. M. Kang and J. Schiefelbein, *Plant Physiology*, 2019, **179**, 1444-1456.
1312 164. C. Seyfferth, J. Renema, J. R. Wendrich, T. Eekhout, R. Seurinck, N. Vandamme, B. Blob, Y. Saeys,
1313 Y. Helariutta, K. D. Birnbaum and B. De Rybel, in *Annual Review of Plant Biology, Vol 72, 2021*,
1314 ed. S. S. Merchant, 2021, vol. 72, pp. 847-866.
1315 165. B. M. Moore, P. Wang, P. Fan, B. Leong, C. A. Schenck, J. P. Lloyd, M. D. Lehti-Shiu, R. L. Last, E.
1316 Pichersky and S.-H. Shiu, *Proceedings of the National Academy of Sciences of the United States*
1317 *of America*, 2019, **116**, 2344-2353.
1318 166. B. M. Moore, P. Wang, P. Fan, A. Lee, B. Leong, Y.-R. Lou, C. A. Schenck, K. Sugimoto, R. Last, M.
1319 D. Lehti-Shiu, C. S. Barry and S.-H. Shiu, *In silico Plants*, 2020, **2**, diaa005.
1320 167. P. Wang, B. M. Moore, S. Uygun, M. D. Lehti-Shiu, C. S. Barry and S.-H. Shiu, *New Phytologist*,
1321 2021, **231**, 475-489.
1322 168. L. P. de Souza, S. Alseekh, F. Scossa and A. R. Fernie, *Nature Methods*, 2021, **18**, 733-746.
1323 169. R. L. Last, A. D. Jones and Y. Shachar-Hill, *Nature Reviews Molecular Cell Biology*, 2007, **8**, 167-
1324 174.
1325 170. H. Horai, M. Arita, S. Kanaya, Y. Nihei, T. Ikeda, K. Suwa, Y. Ojima, K. Tanaka, S. Tanaka, K.
1326 Aoshima, Y. Oda, Y. Kakazu, M. Kusano, T. Tohge, F. Matsuda, Y. Sawada, M. Y. Hirai, H.
1327 Nakanishi, K. Ikeda, N. Akimoto, T. Maoka, H. Takahashi, T. Ara, N. Sakurai, H. Suzuki, D. Shibata,
1328 S. Neumann, T. Iida, K. Tanaka, K. Funatsu, F. Matsuura, T. Soga, R. Taguchi, K. Saito and T.
1329 Nishioka, *Journal of Mass Spectrometry*, 2010, **45**, 703-714.
1330 171. H. Tsugawa, R. Nakabayashi, T. Mori, Y. Yamada, M. Takahashi, A. Rai, R. Sugiyama, H.
1331 Yamamoto, T. Nakaya, M. Yamazaki, R. Kooke, J. A. Bac-Molenaar, N. Oztolan-Erol, J. J. B.
1332 Keurentjes, M. Arita and K. Saito, *Nature Methods*, 2019, **16**, 295-298.
1333 172. M. Wang, J. J. Carver, V. V. Phelan, L. M. Sanchez, N. Garg, Y. Peng, N. Don Duy, J. Watrous, C. A.
1334 Kapono, T. Luzzatto-Knaan, C. Porto, A. Bouslimani, A. V. Melnik, M. J. Meehan, W. T. Liu, M.
1335 Criisemann, P. D. Boudreau, E. Esquenazi, M. Sandoval-Calderon, R. D. Kersten, L. A. Pace, R. A.
1336 Quinn, K. R. Duncan, C.-C. Hsu, D. J. Floros, R. G. Gavilan, K. Kleigrewe, T. Northen, R. J. Dutton,
1337 D. Parrot, E. E. Carlson, B. Aigle, C. F. Michelsen, L. Jelsbak, C. Sohlenkamp, P. Pevzner, A.
1338 Edlund, J. McLean, J. Piel, B. T. Murphy, L. Gerwick, C.-C. Liaw, Y.-L. Yang, H.-U. Humpf, M.
1339 Maansson, R. A. Keyzers, A. C. Sims, A. R. Johnson, A. M. Sidebottom, B. E. Sedio, A. Klitgaard, C.
1340 B. Larson, C. A. Boya P, D. Torres-Mendoza, D. J. Gonzalez, D. B. Silva, L. M. Marques, D. P.
1341 Demarque, E. Pociute, E. C. O'Neill, E. Briand, E. J. N. Helfrich, E. A. Granatosky, E. Glukhov, F.
1342 Ryffel, H. Houson, H. Mohimani, J. J. Kharbush, Y. Zeng, J. A. Vorholt, K. L. Kurita, P. Charusanti,
1343 K. L. McPhail, K. F. Nielsen, L. Vuong, M. Elfeki, M. F. Traxler, N. Engene, N. Koyama, O. B. Vining,
1344 R. Baric, R. R. Silva, S. J. Mascuch, S. Tomasi, S. Jenkins, V. Macherla, T. Hoffman, V. Agarwal, P.

G. Williams, J. Dai, R. Neupane, J. Gurr, A. M. C. Rodríguez, A. Lamsa, C. Zhang, K. Dorrestein, B. M. Duggan, J. Almaliti, P.-M. Allard, P. Phapale, L.-F. Nothias, T. Alexandrovr, M. Litaudon, J.-L. Wolfender, J. E. Kyle, T. O. Metz, T. Peryea, N. Dac-Trung, D. VanLeer, P. Shinn, A. Jadhav, R. Müller, K. M. Waters, W. Shi, X. Liu, L. Zhang, R. Knight, P. R. Jensen, B. Ø. Palsson, K. Pogliano, R. G. Linington, M. Gutiérrez, N. P. Lopes, W. H. Gerwick, B. S. Moore, P. C. Dorrestein and N. Bandeira, *Nature Biotechnology*, 2016, **34**, 828-837.

173. L. W. Sumner, Z. Lei, B. J. Nikolau and K. Saito, *Natural Product Reports*, 2015, **32**, 212-229.

174. Y. Dong, P. Sonawane, H. Cohen, G. Polturak, L. Feldberg, S. H. Avivi, I. Rogachev and A. Aharoni, *New Phytologist*, 2020, **228**, 1986-2002.

175. S. Alseekh, I. Ofner, Z. Liu, S. Osorio, J. Vallarino, R. L. Last, D. Zamir, T. Tohge and A. R. Fernie, *The Plant Journal*, 2020, **103**, 2007-2024.

176. C. N. Hirsch, C. D. Hirsch, K. Felcher, J. Coombs, D. Zarka, A. Van Deynze, W. De Jong, R. E. Veilleux, S. Jansky, P. Bethke, D. S. Douches and C. R. Buell, *G3-Genes Genomes Genetics*, 2013, **3**, 1003-1013.

177. P. Gramazio, J. Prohens, M. Plazas, G. Mangino, F. J. Herraiz and S. Vilanova, *Frontiers in Plant Science*, 2017, **8**, Article number 1477.

178. M. Sulli, L. Barchi, L. Toppino, G. Diretto, T. Sala, S. Lanteri, G. L. Rotino and G. Giuliano, *Frontiers in Plant Science*, 2021, **12**, Article number 638195.

179. A. V. Levina, O. Hoekenga, M. Gordin, C. Broeckling and W. S. De Jong, *Crop Science*, 2021, **61**, 591-603.

180. V. Tzin, S. Malitsky, M. M. Ben Zvi, M. Bedair, L. Sumner, A. Aharoni and G. Galili, *New Phytologist*, 2012, **194**, 430-439.

181. Y. Zhang, E. Butelli, S. Alseekh, T. Tohge, G. Rallapalli, J. Luo, P. G. Kawar, L. Hill, A. Santino, A. R. Fernie and C. Martin, *Nature Communications*, 2015, **6**, Article number 8635.

182. E. Lewinsohn and M. Gijzen, *Plant Science*, 2009, **176**, 161-169.

183. C. G. Jones, M. W. Martynowycz, J. Hattne, T. J. Fulton, B. M. Stoltz, J. A. Rodriguez, H. M. Nelson and T. Gonen, *ACS Central Science*, 2018, **4**, 1587-1592.

184. T. Gruene, J. T. C. Wennmacher, C. Zaubitzer, J. J. Holstein, J. Heidler, A. Fecteau-Lefebvre, S. De Carlo, E. Müller, K. N. Goldie, I. Regeni, T. Li, G. Santiso-Quinones, G. Steinfeld, S. Handschin, E. van Genderen, J. A. van Bokhoven, G. H. Clever and R. Pantelic, *Angewandte Chemie-International Edition*, 2018, **57**, 16313-16317.

185. L. J. Kim, M. Ohashi, Z. Zhang, D. Tan, M. Asay, D. Cascio, J. A. Rodriguez, Y. Tang and H. M. Nelson, *Nature Chemical Biology*, 2021, **17**, 872-877.

186. J. Reed, M. J. Stephenson, K. Miettinen, B. Brouwer, A. Leveau, P. Brett, R. J. M. Goss, A. Goossens, M. A. O'Connell and A. Osbourn, *Metabolic Engineering*, 2017, **42**, 185-193.

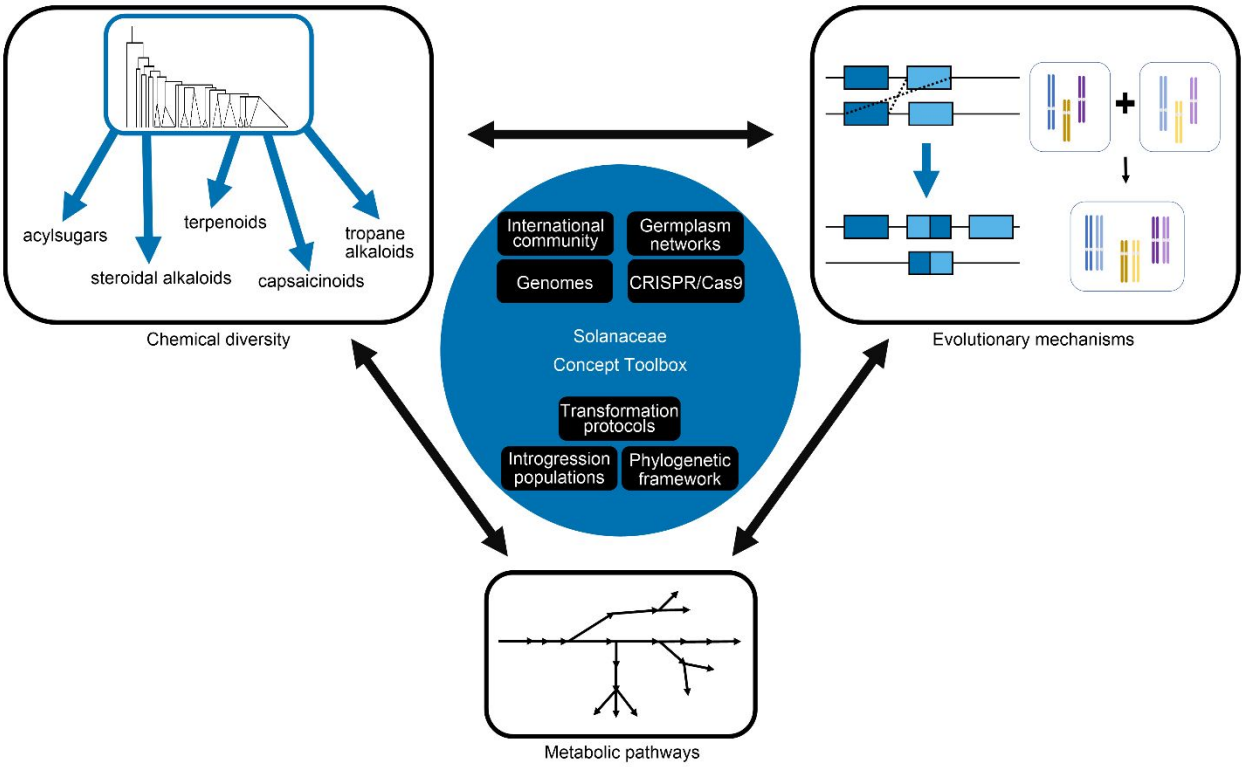


Figure 1. Solanaceae as a model family for specialized metabolism evolution studies. The Solanaceae concept toolbox connects biodiversity, genetics, and evolutionary mechanisms to each other. Chemical diversity informs metabolic pathway discovery, which in turn reveals evolutionary mechanisms underlying chemical diversity.

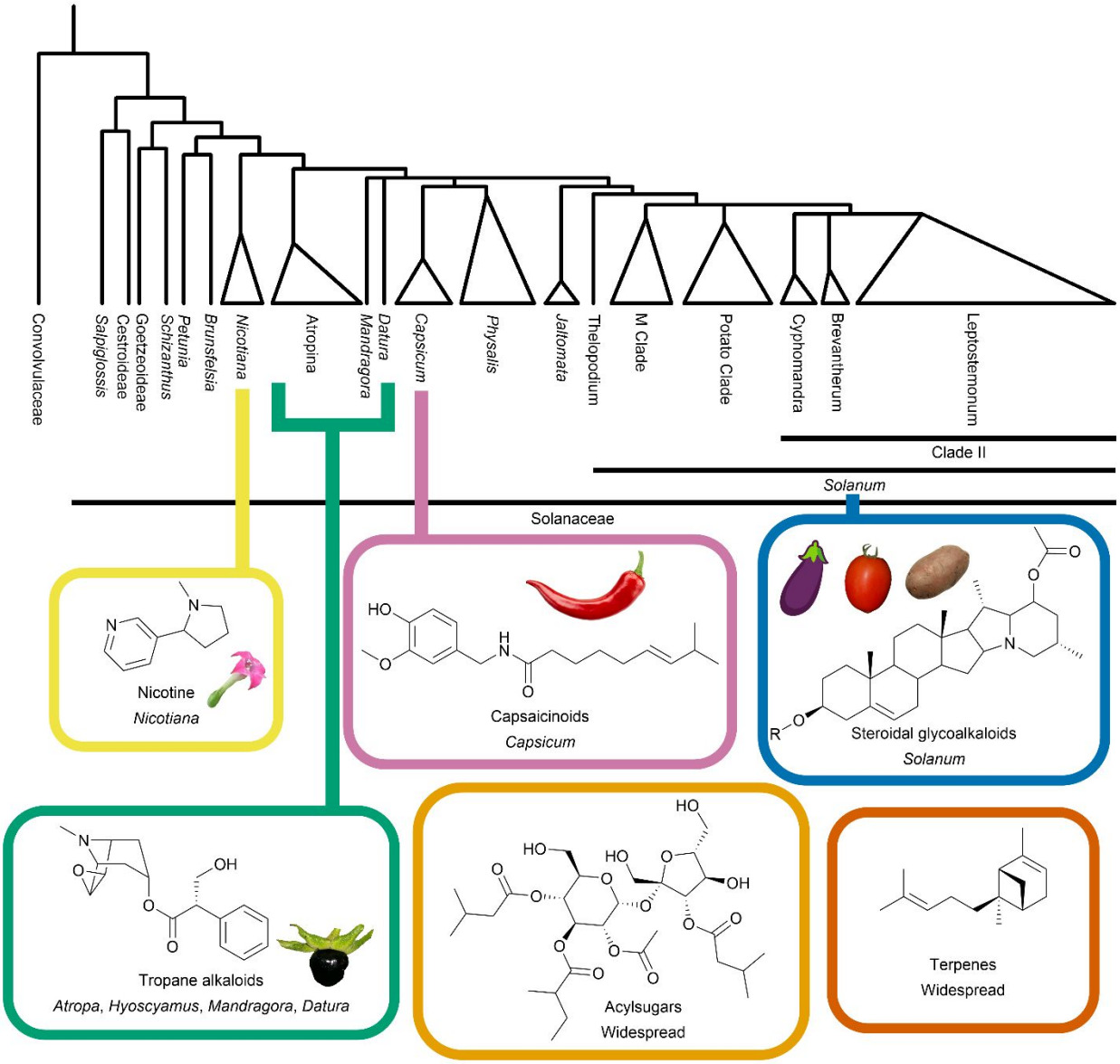


Figure 2. Phylogenetic distribution of major Solanaceae specialized metabolite classes. The Solanaceae family produces specialized metabolites of multiple chemical classes. A simplified phylogeny of the Solanaceae family is shown based on prior determination of phylogenetic relationships[11, 12]. Major metabolite classes are mapped to the corresponding clades that produce high amounts of those metabolites and / or act as model species for studying their biosynthesis and evolution. Metabolites may not be distributed solely in the noted phylogenetic group. Additional information on metabolite distribution is provided throughout the text of this article.

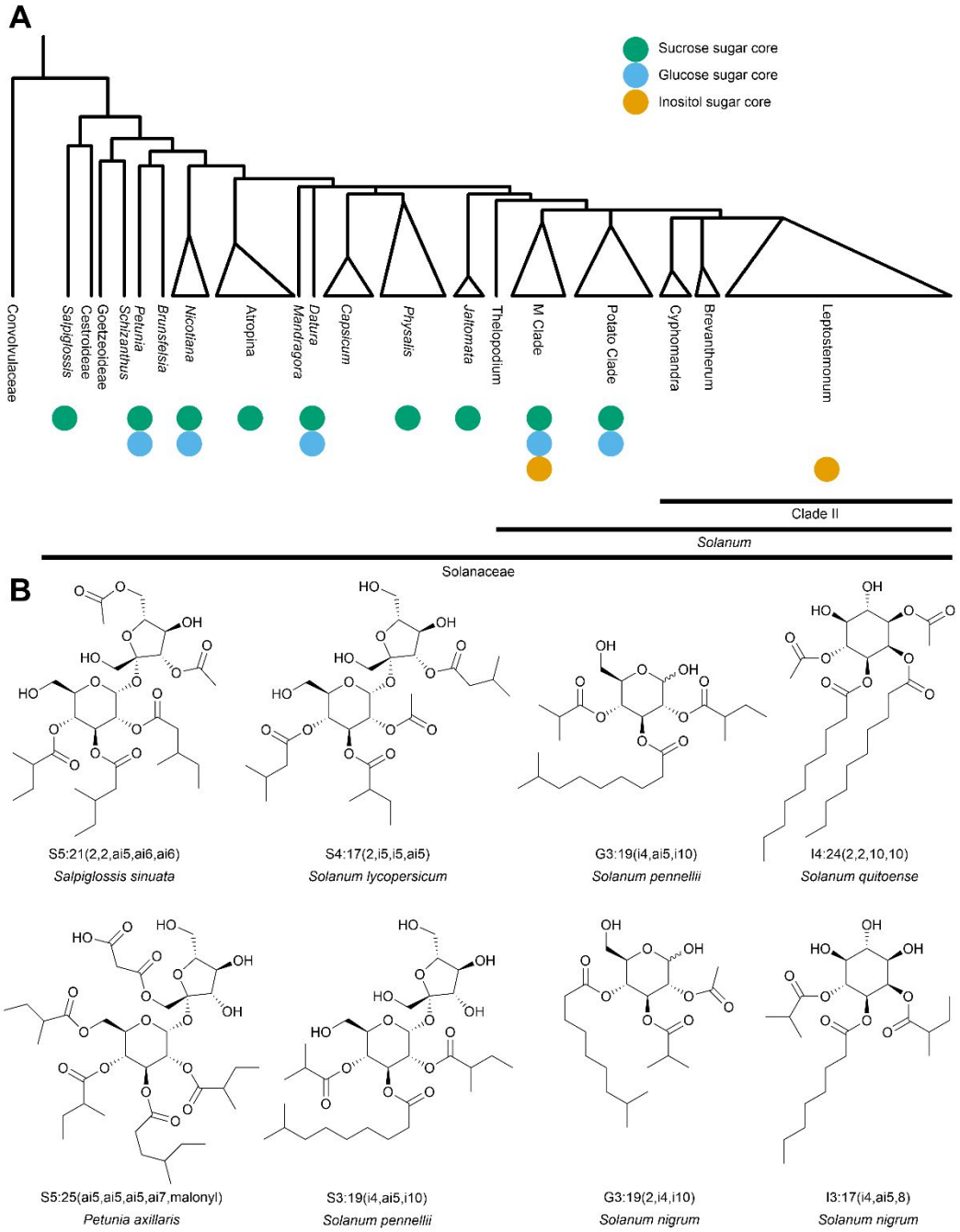


Figure 3. Phylogenetic distribution of acylsugar core types. (A) Simplified Solanaceae phylogeny with acylsugar core type placed on each lineage with characterized acylsugars. The phylogenetic tree is based upon previously published Solanaceae and *Solanum* trees[11, 12]. (B) Characteristic acylsugar structures produced by Solanaceae species[36, 37, 49, 50, 53, 57, 72-75]. Acylsugar nomenclature is given for each compound where the first letter represents the sugar core (S for sucrose, G for glucose, I for inositol); the first number represents the number of acylations; the number after the colon represents the number of carbons in acyl chains; and the individual acyl chains are listed inside parentheses (ai = anteiso, i = iso).

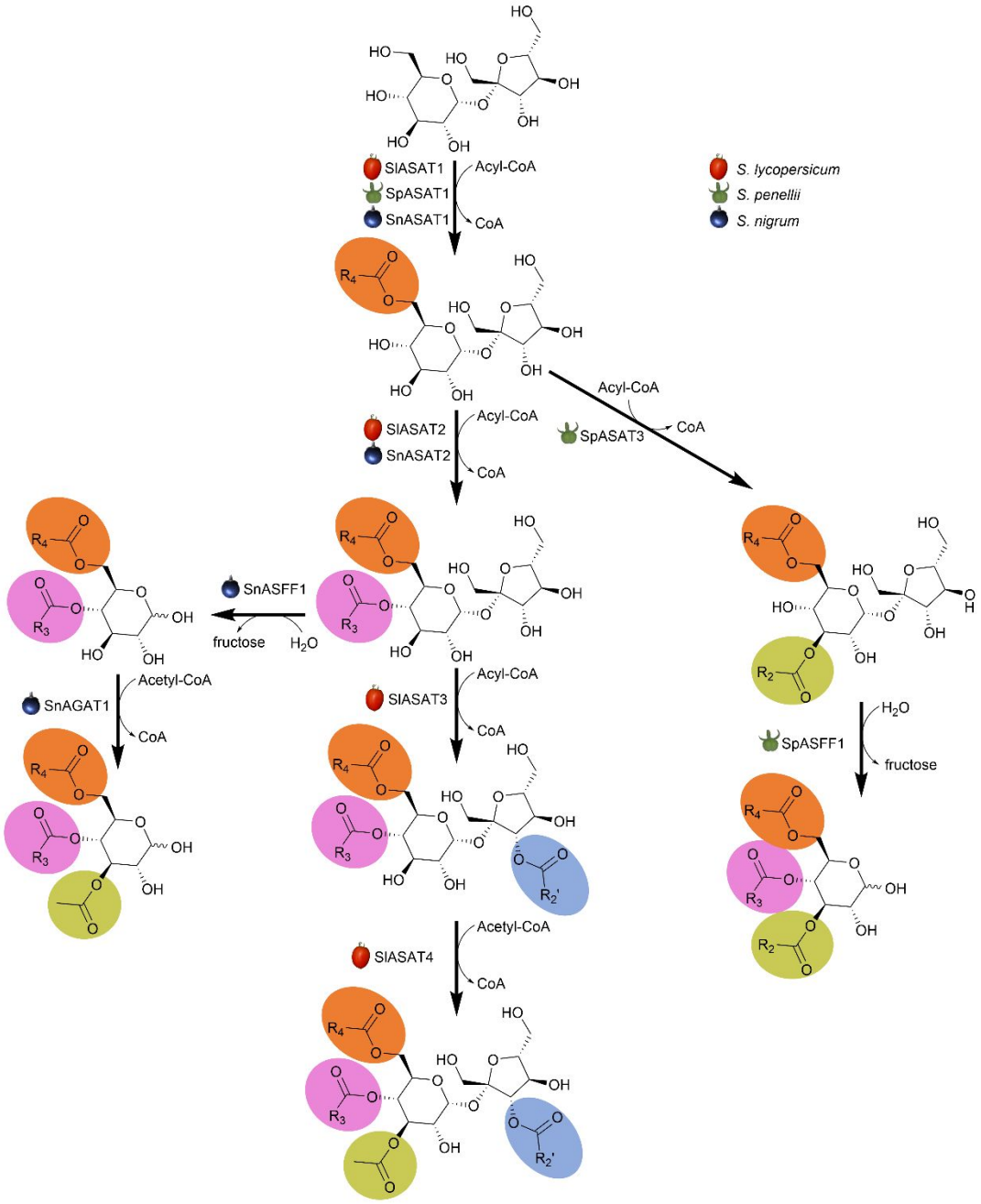


Figure 4. Acylsucrose and acylglucose pathway diversity in *Solanum* species. The acylsucrose and acylglucose biosynthesis pathways for *S. nigrum*, *S. lycopersicum* and *S. pennellii*. All three biosynthetic pathways begin by acylating sucrose[24, 63, 64, 68, 72]. Sequential acylations produce tetraacylsucroses, triacylsucroses, and diacylsucroses for *S. lycopersicum*, *S. pennellii*, and *S. nigrum*, respectively. *S. pennellii* triacylsucroses and *S. nigrum* diacylsucroses are cleaved by ASFF enzymes to form triacylglucoses and diacylglucoses, respectively[68, 72]. *S. nigrum* diacylglucose is acetylated by SnAGAT1 to form a triacylglucose[72]. ASAT, acylsucrose acyltransferase; AGAT, acylglucose acyltransferase; ASFF, acylsugar fructofuranosidase; CoA, CoenzymeA.

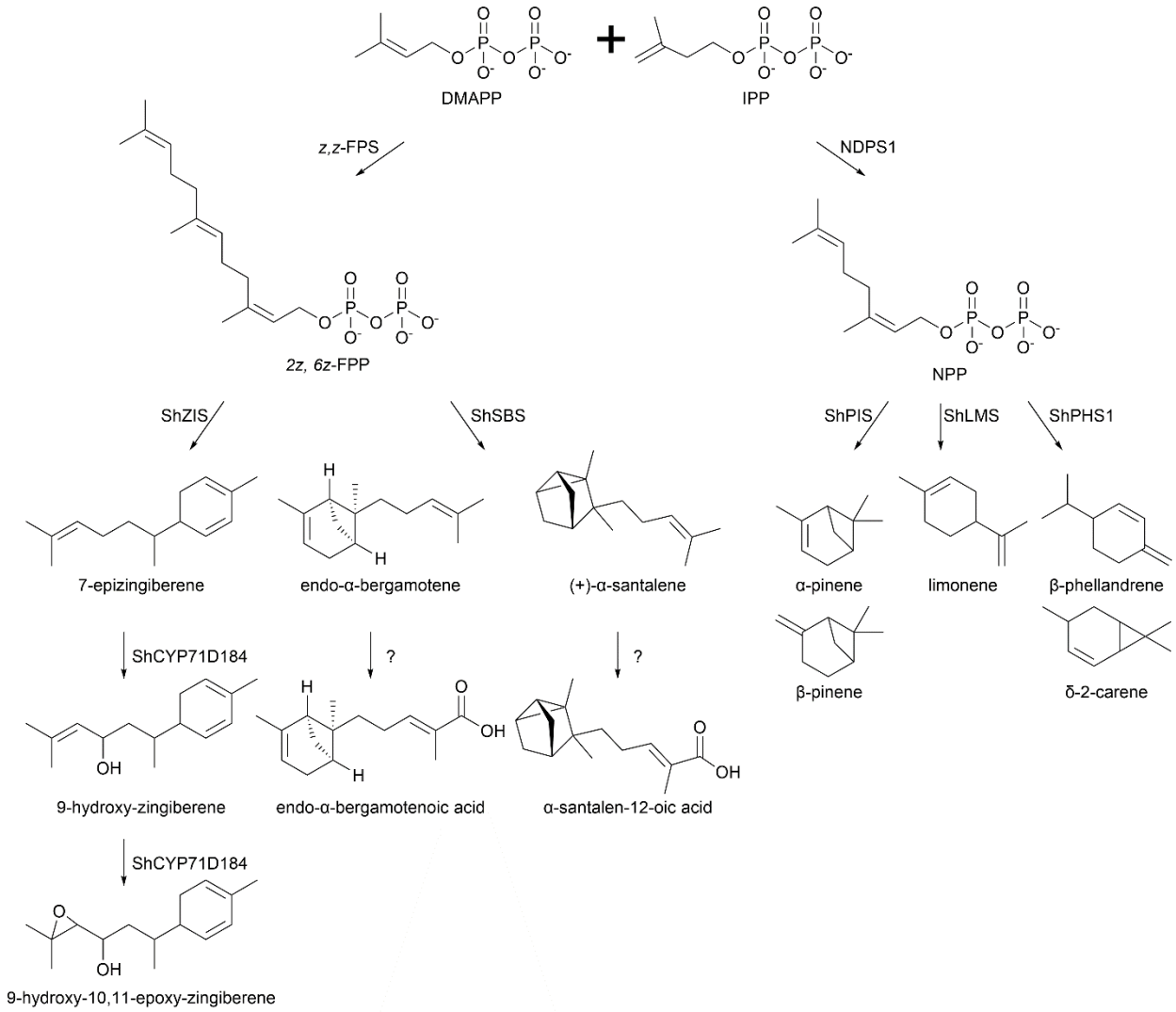
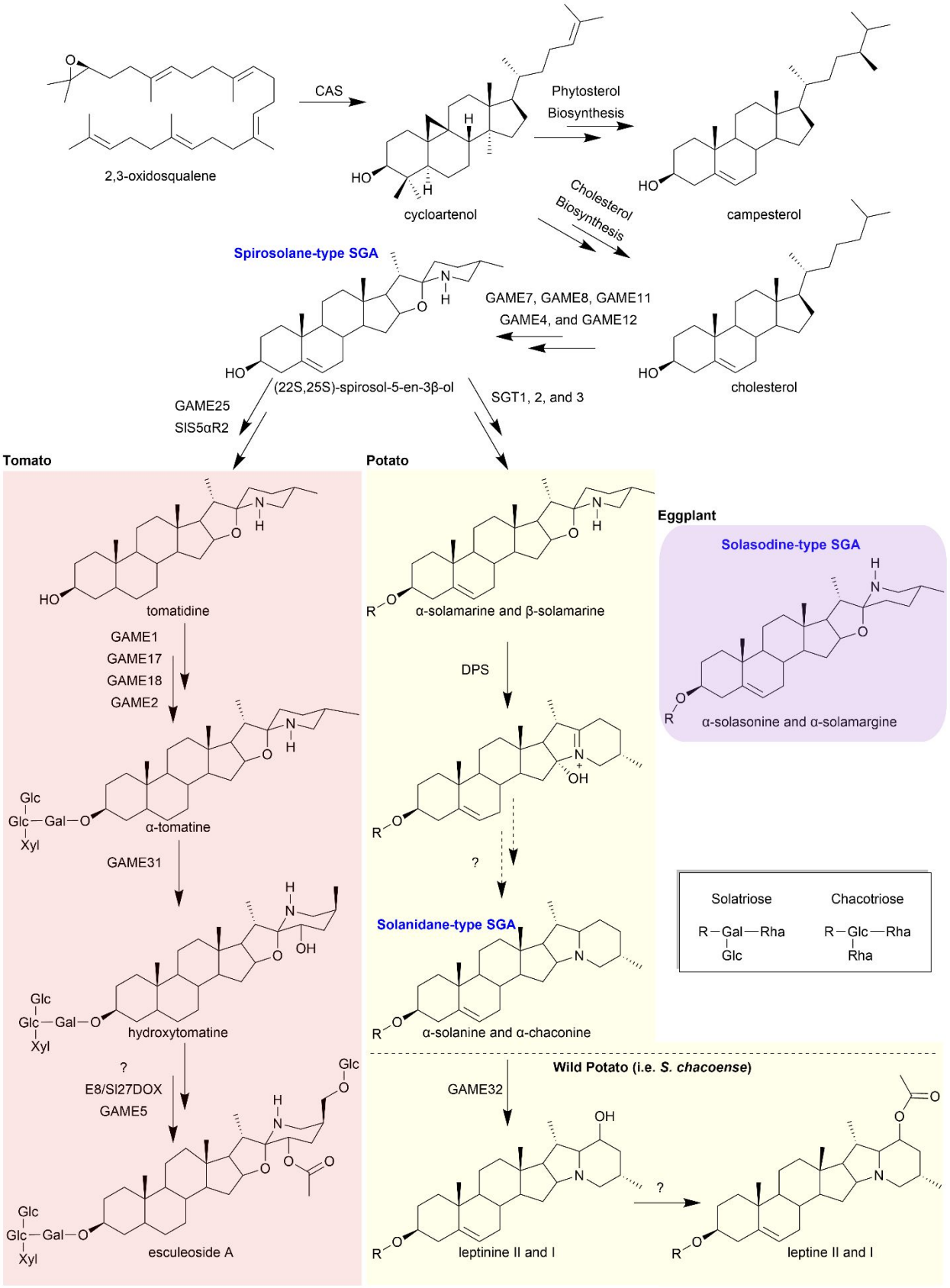


Figure 5. Terpenoid biosynthesis in the trichomes of *Solanum habrochaites* derived from *cisoid* substrates. NDPS1 catalyzes the condensation of a single molecule of DMAPP and IPP to form NPP (C10)[88]. In contrast, *z,z*-FPS catalyzes the formation of *2z,6z*-FPP (C15) through sequential condensation of two molecules of IPP with a single molecule of DMAPP[89]. In distinct NPP producing accessions of *S. habrochaites* the monoterpene synthases, ShPIS, ShLMS, and ShPHS1 catalyze the cyclization of NPP to form monoterpenes[87]. In a subset of *2z,6z*-FPP forming accessions, the sesquiterpene synthase, ShSBS catalyzes the formation of endo-α-bergamotene and (+)-α-santalene[87, 89]. These sesquiterpenes are converted to their corresponding acids by unknown enzymes. In a distinct subset of *2z,6z*-FPP producing accessions, ShZIS catalyzes the formation of 7-epizingiberene, which is sequentially oxidized by ShCYP71D184 to 9-hydroxy-zingiberene and 9-hydroxy-10, 11-epoxy-zingiberene[87, 92, 95]. In trichomes of cultivated tomato, *S. lycopersicum*, only orthologs of NDPS1 and ShPHS1 are present resulting in the formation of β-phellandrene and δ-2-carene[88]. Thus, *cisoid* substrate derived terpene diversity is attenuated in *S. lycopersicum* in comparison to *S. habrochaites*. Abbreviations are as follows: DMAPP, dimethylallyl diphosphate; IPP, isopentenyl diphosphate; NPP, neryl diphosphate; *2z,6z*-FPP, *2z,6z*-farnesyl diphosphate; ShZIS, zingiberene synthase; ShSBS, santalene and bergamotene synthase; ShPIS, pinene synthase; ShLMS, limonene synthase; ShPHS1, β-phellandrene synthase.


2,3-oxidosqualene
CAS
cycloartenol
Phytosterol Biosynthesis
campesterol
Cholesterol Biosynthesis
Spirosolane-type SGA
GAME7, GAME8, GAME11
GAME4, and GAME12
(22S,25S)-spirosol-5-en-3β-ol
cholesterol
GAME25
SlS5αR2
SGT1, 2, and 3
Tomato
Potato
Eggplant
Solasodine-type SGA
tomatidine
α-solamarine and β-solamarine
α-solasonine and α-solamargine
GAME1
GAME17
GAME18
GAME2
DPS
α-tomatine
GAME31
Solatriose
Chacotriose
hydroxytomatine
Solanidane-type SGA
α-solanine and α-chaconine
E8/Sl27DOX
GAME5
Wild Potato (i.e. S. chacoense)
GAME32
esculeoside A
leptinine II and I
leptine II and I

1434

1435 **Figure 6. Steroidal glycoalkaloid biosynthesis in *Solanum*.** CAS cyclizes 2,3-oxidosqualene from the
1436 mevalonate pathway to form cycloartenol a common metabolite in both phytosterol and cholesterol
1437 biosynthesis. Cycloartenol is converted to campesterol by a ten-step pathway and through a nine-step
1438 pathway to form cholesterol[121]. Following the production of cholesterol, five GAME enzymes are required
1439 to produce the spirosolane-type SGA core[8]. In tomato (red shaded box), GAME25 catalyzes the first of
1440 four steps resulting in tomatidine formation via the reduction of the spirosolane-type SGA core[123, 124].
1441 Subsequent sugar additions by GAME1, GAME17, GAME18, and GAME2 result in the formation of α-
1442 tomatine[8]. GAME31, E8/Sl27DOX, GAME5, and an unknown acetyltransferase catalyze the fruit ripening
1443 associated formation of esculeoside A from α-tomatine[117, 126-129]. In potato (yellow shading), the addition
1444 of solatriose and chacotriose moieties by sequential sugar additions to (22S, 25S)-spirosol-5-en-3β-ol
1445 results in the formation of α- and β-solamarine, respectively[10]. The oxidization of α- and β-solamarine by
1446 DPS represents the first step in α-solanine and α-chaconine, Solanidane-type SGA, formation[10]. In *S.*
1447 *chacoense*, α-solanine and α-chaconine are oxidized by GAME32 to form leptinines, and leptine formation
1448 requires the acetylation at the GAME32 introduced oxidation[117]. The solasodine-type SGAs (α-solasonine
1449 and α-solamargine) are the main SGAs in eggplant (purple shading) and contain solatriose and chacotriose
1450 moieties at the C-3 position, respectively. The biosynthetic mechanism leading to the stereochemical
1451 difference in spirosolane and solasodine cores remains uncharacterized[10, 120]. Enzyme abbreviations are
1452 as follows: CAS, cycloartenol synthase; GAME, glycoalkaloid metabolism; SlS5αR2, steroid 5α-reductase
1453 2; SGT, solanidine glycosyltransferase; DPS, dioxygenase for potato solanidane synthesis; E8/Sl27DOX, α-
1454 tomatine 27-hydroxylase; Gal, galactose; Glc, glucose; Xyl, xylose; Rha, Rhamnose.

1455

1456

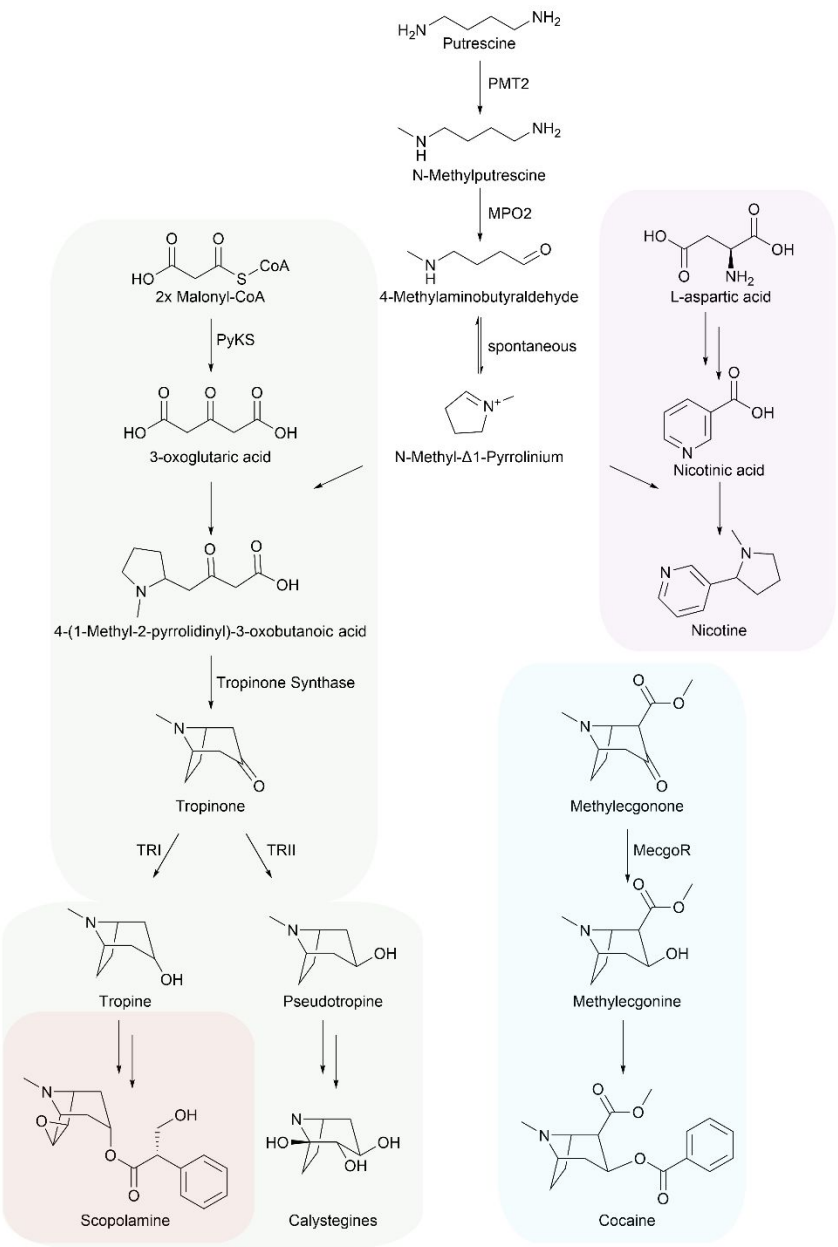


Figure 7. Evolutionary trajectories of tropane and nicotine formation in distinct plant lineages. Comparison of tropane and nicotine alkaloid biosynthesis reveals examples of both convergent (cocaine biosynthesis in *E.* coca) and divergent (nicotine biosynthesis) evolution[132, 142]. Scopolamine (orange) and nicotine (purple) represent alternative fates of the *N*-methylpyrrolinium cation in different genera of the Solanaceae. The use of an aldo-keto reductase enzyme (MecgoR) in the penultimate step of cocaine biosynthesis (blue) contrasts with catalysis by short-chain dehydrogenase/reductase (SDR) family enzymes (TRI and TRII) in scopolamine formation (green)[132]. *Not shown is catalysis by a single, bifunctional SDR to produce both tropine and pseudotropine in Brassicaceae[131]. Tropanol biosynthesis (green) is widely distributed across the Solanaceae compared to the biosynthesis of tropane aromatic esters such as scopolamine (orange)[137]. Enzyme abbreviations are as follows: PMT2, Putrescine *N*-methyltransferase 2; MPO2, *N*-methylputrescine oxidase 2; PyKS, Polyketide Synthase; TRI, Tropinone reductase I; TRII, Tropinone Reductase II; MecgoR, Methylecgonone reductase.

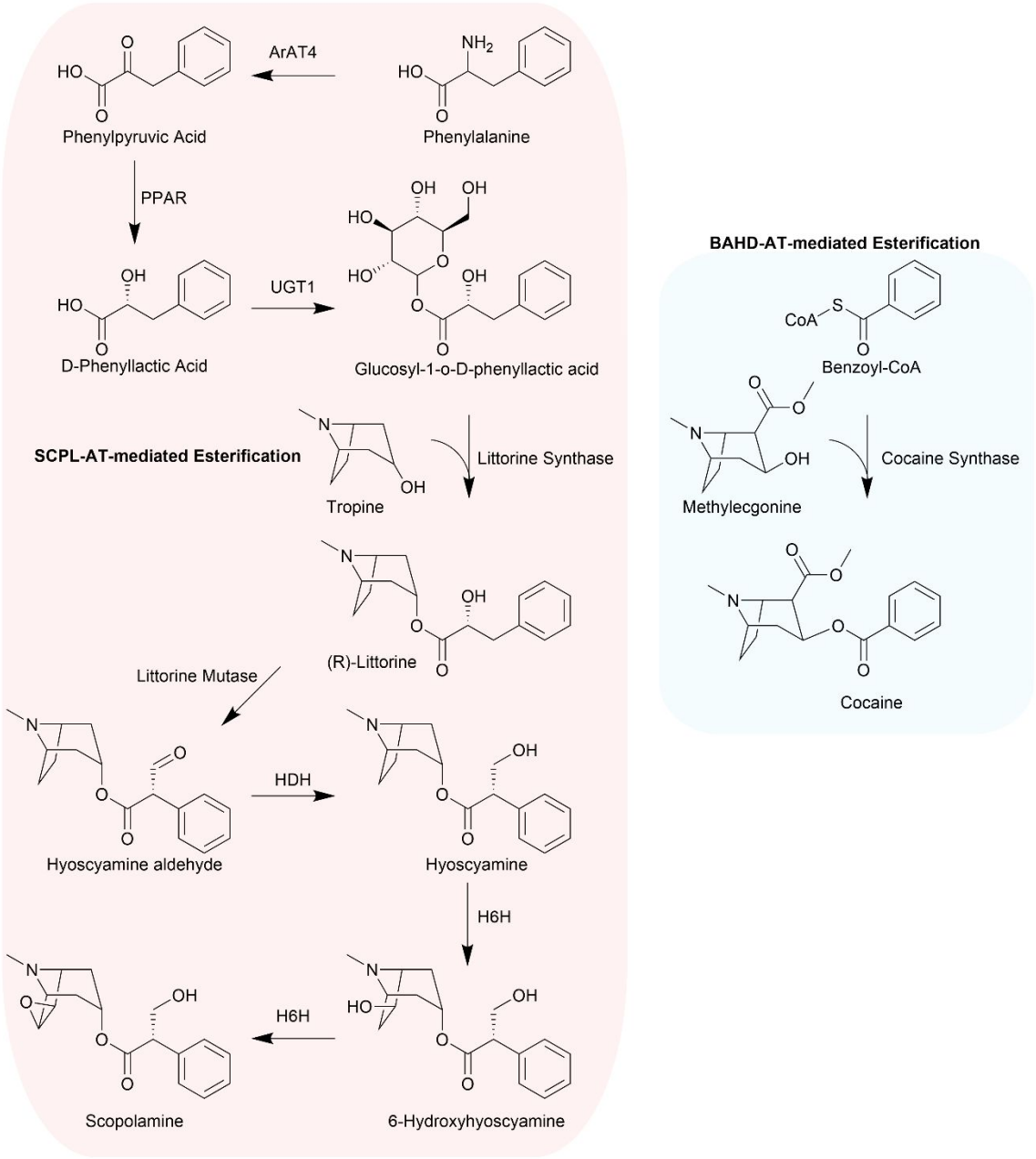


Figure 8. Independent evolution of tropane aromatic ester formation in Solanaceae and Erythroxylaceae. Scopolamine biosynthesis requires the biosynthesis of D-phenyllactic acid via a two-step process mediated by ArAT4 and PPAR[147, 148]. D-phenyllactic acid is glycosylated by UGT1 to form a glucose ester of phenyllactic acid, which is used, along with tropine, as substrate for littorine biosynthesis by Littorine Synthase, a serine carboxypeptidase-like acyltransferase[136]. Three enzymes, Littorine Mutase, HDH, and H6H, are required for the conversion of littorine to scopolamine[138, 150, 151]. In contrast, cocaine biosynthesis utilizes a BAHD acyl-transferase and coenzyme A donor to facilitate the transfer of a benzoyl moiety on to methylecgonine, the *E. coca* tropanol, to form cocaine[153]. Enzyme abbreviations are as follows: ArAT4, Aromatic amino acid transferase 4; PPAR, phenylpyruvic acid reductase; UGT1, UDP-glycosyltransferase 1; HDH, Hyoscyamine dehydrogenase; H6H, hyoscyamine-6-hydroxylase.

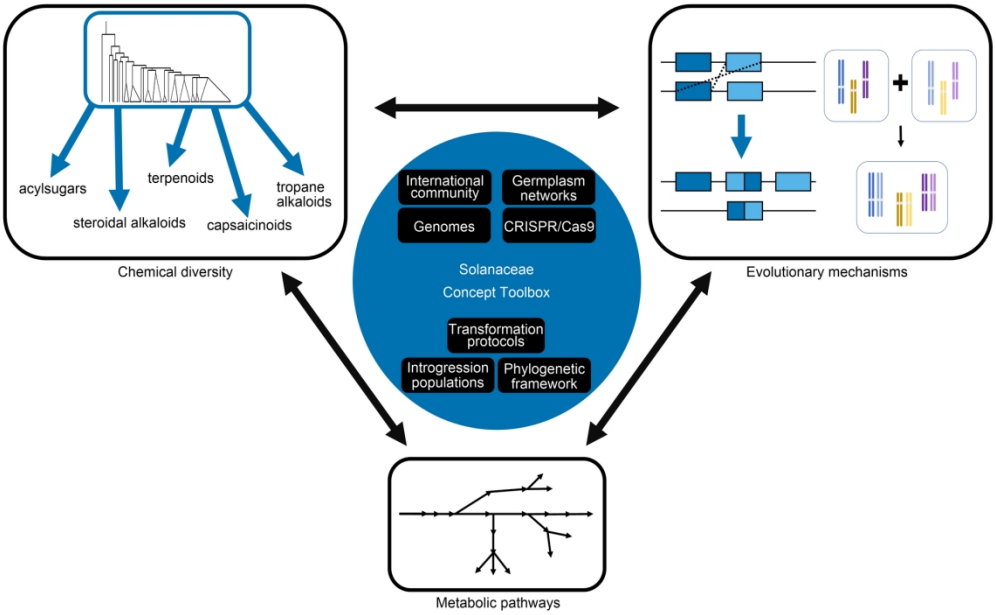


Figure 1. Solanaceae as a model family for specialized metabolism evolution studies. The Solanaceae concept toolbox connects biodiversity, genetics, and evolutionary mechanisms to each other. Chemical diversity informs metabolic pathway discovery, which in turn reveals evolutionary mechanisms underlying chemical diversity.

173x107mm (300 x 300 DPI)

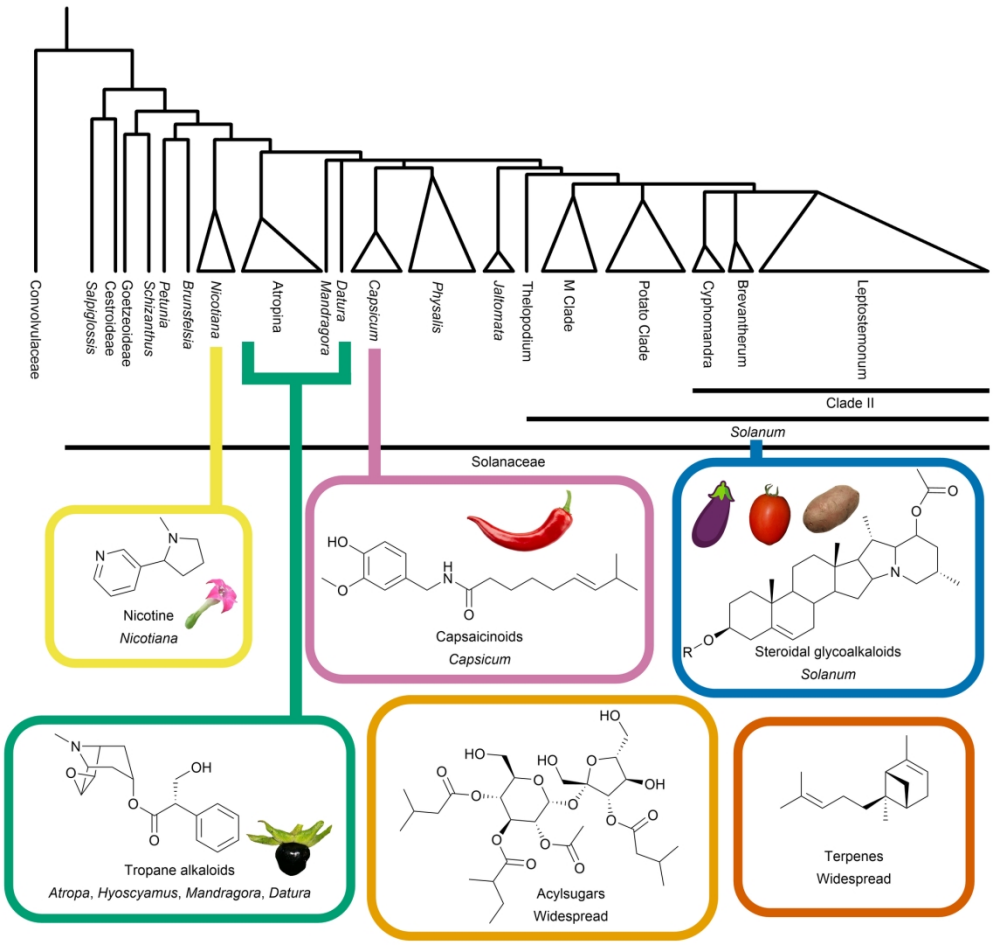


Figure 2. Phylogenetic distribution of major Solanaceae specialized metabolite classes. The Solanaceae family produces specialized metabolites of multiple chemical classes. A simplified phylogeny of the Solanaceae family is shown based on prior determination of phylogenetic relationships[11, 12]. Major metabolite classes are mapped to the corresponding clades that produce high amounts of those metabolites and / or act as model species for studying their biosynthesis and evolution. Metabolites may not be distributed solely in the noted phylogenetic group. Additional information on metabolite distribution is provided throughout the text of this article.

176x167mm (300 x 300 DPI)

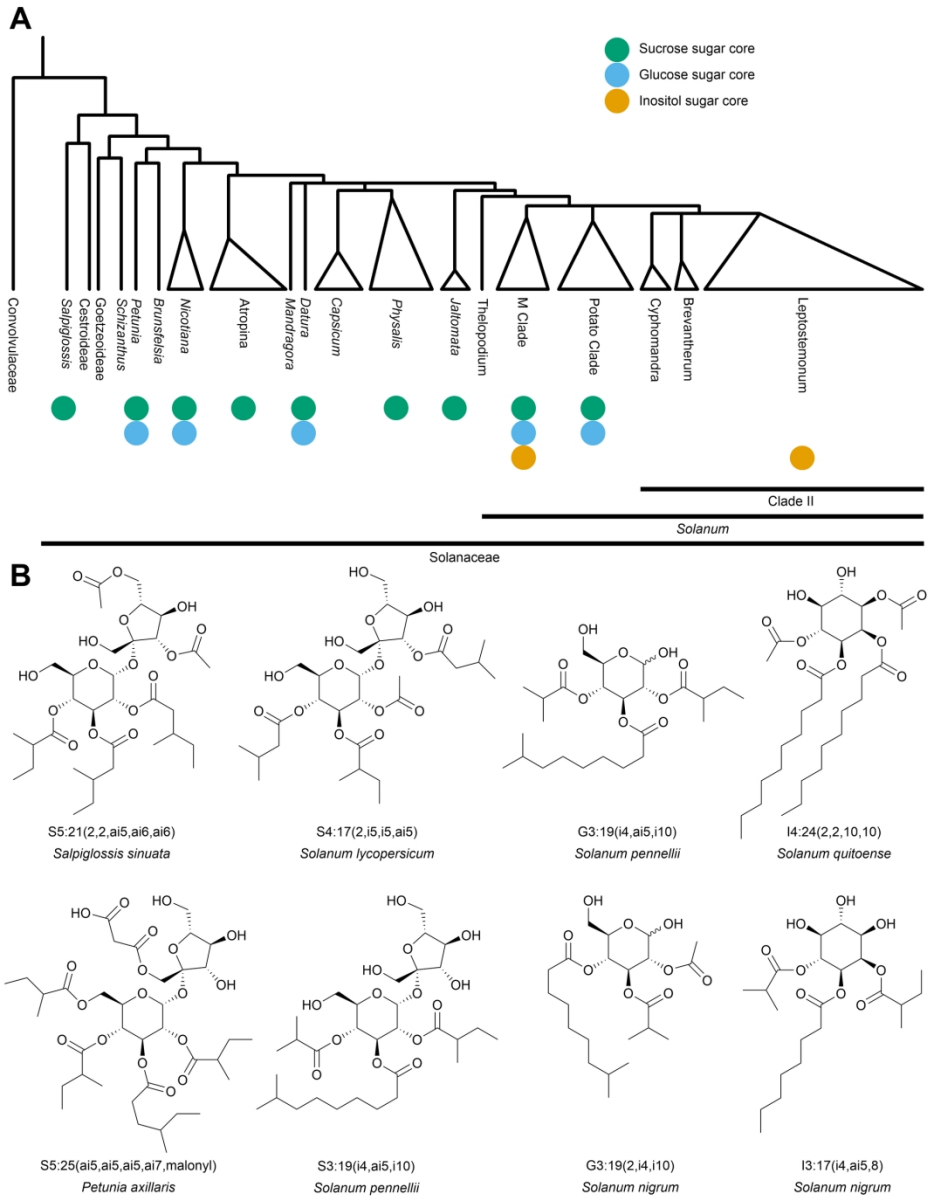


Figure 3. Phylogenetic distribution of acylsugar core types. (A) Simplified Solanaceae phylogeny with acylsugar core type placed on each lineage with characterized acylsugars. The phylogenetic tree is based upon previously published Solanaceae and Solanum trees[11, 12]. (B) Characteristic acylsugar structures produced by Solanaceae species[36, 37, 49, 50, 53, 57, 72-75]. Acylsugar nomenclature is given for each compound where the first letter represents the sugar core (S for sucrose, G for glucose, I for inositol); the first number represents the number of acylations; the number after the colon represents the number of carbons in acyl chains; and the individual acyl chains are listed inside parentheses (ai = anteiso, i = iso).

172x221mm (300 x 300 DPI)

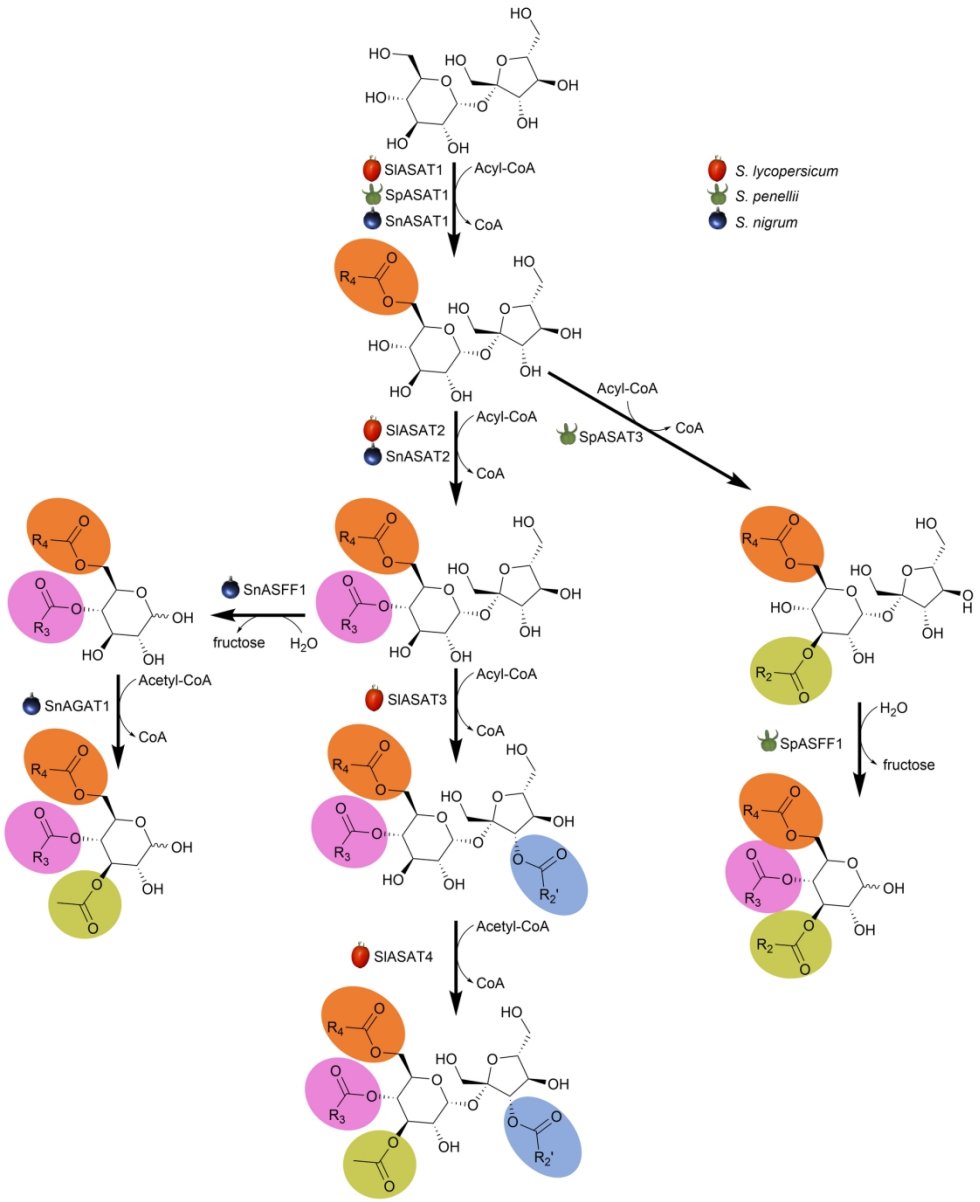


Figure 4. Acylsucrose and acylglucose pathway diversity in *Solanum* species. The acylsucrose and acylglucose biosynthesis pathways for *S. nigrum*, *S. lycopersicum* and *S. pennellii*. All three biosynthetic pathways begin by acylating sucrose[24, 63, 64, 68, 72]. Sequential acylations produce tetraacylsucroses, triacylsucroses, and diacylsucroses for *S. lycopersicum*, *S. pennellii*, and *S. nigrum*, respectively. *S. pennellii* triacylsucroses and *S. nigrum* diacylsucroses are cleaved by ASFF enzymes to form triacylglucoses and diacylglucoses, respectively[68, 72]. *S. nigrum* diacylglucose is acetylated by SnAGAT1 to form a triacylglucose[72]. ASAT, acylsucrose acyltransferase; AGAT, acylglucose acyltransferase; ASFF, acylsugar fructofuranosidase; CoA, CoenzymeA.

167x202mm (300 x 300 DPI)

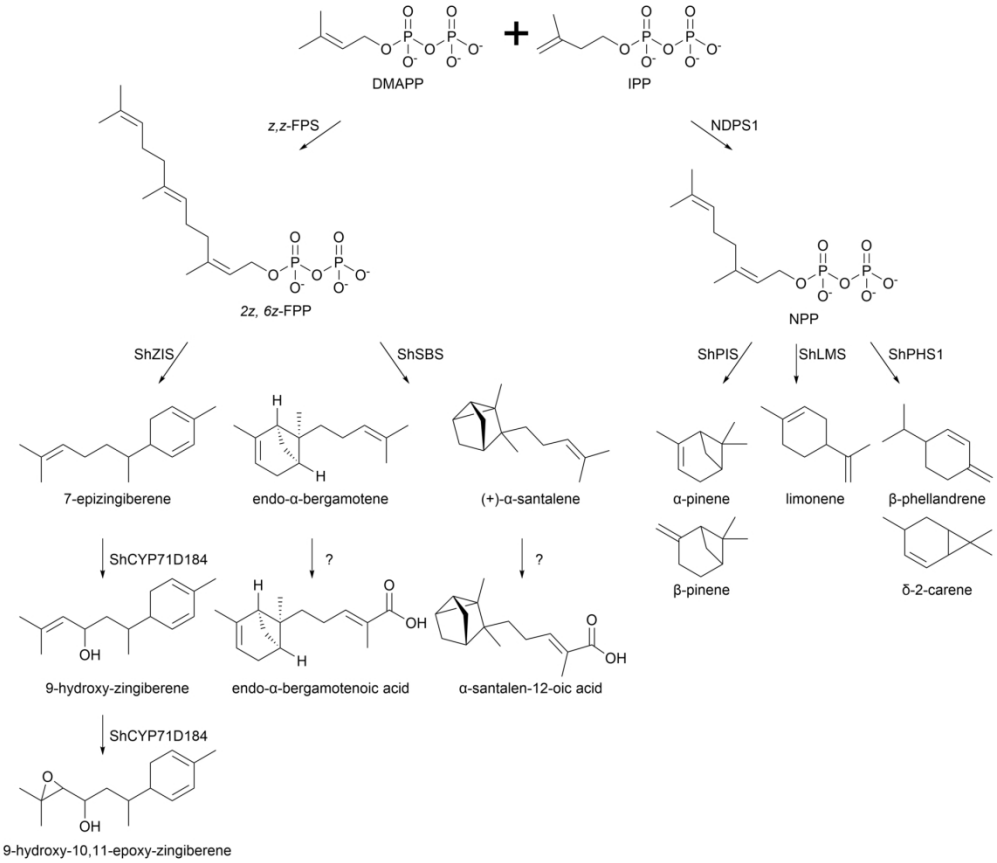


Figure 5. Terpenoid biosynthesis in the trichomes of *Solanum habrochaites* derived from cisoid substrates. NDPS1 catalyzes the condensation of a single molecule of DMAPP and IPP to form NPP (C10)[88]. In contrast, *z*,*z*-FPS catalyzes the formation of 2*z*,6*z*-FPP (C15) through sequential condensation of two molecules of IPP with a single molecule of DMAPP[89]. In distinct NPP producing accessions of *S. habrochaites* the monoterpene synthases, ShPIS, ShLMS, and ShPHS1 catalyze the cyclization of NPP to form monoterpenes[87]. In a subset of 2*z*,6*z*-FPP forming accessions, the sesquiterpene synthase, ShSBS catalyzes the formation of endo-α-bergamotene and (+)-α-santalene[87, 89]. These sesquiterpenes are converted to their corresponding acids by unknown enzymes. In a distinct subset of 2*z*,6*z*-FPP producing accessions, ShZIS catalyzes the formation of 7-epizingiberene, which is sequentially oxidized by ShCYP71D184 to 9-hydroxy-zingiberene and 9-hydroxy-10, 11-epoxy-zingiberene[87, 92, 95]. In trichomes of cultivated tomato, *S. lycopersicum*, only orthologs of NDPS1 and ShPHS1 are present resulting in the formation of β-phellandrene and δ-2-carene[88]. Thus, cisoid substrate derived terpene diversity is attenuated in *S. lycopersicum* in comparison to *S. habrochaites*. Abbreviations are as follows: DMAPP, dimethylallyl diphosphate; IPP, isopentenyl diphosphate; NPP, neryl diphosphate; 2*z*,6*z*-FPP, 2*z*,6*z*-farnesyl diphosphate; ShZIS, zingiberene synthase; ShSBS, santalene and bergamotene synthase; ShPIS, pinene synthase; ShLMS, limonene synthase; ShPHS1, β-phellandrene synthase.

169x146mm (300 x 300 DPI)

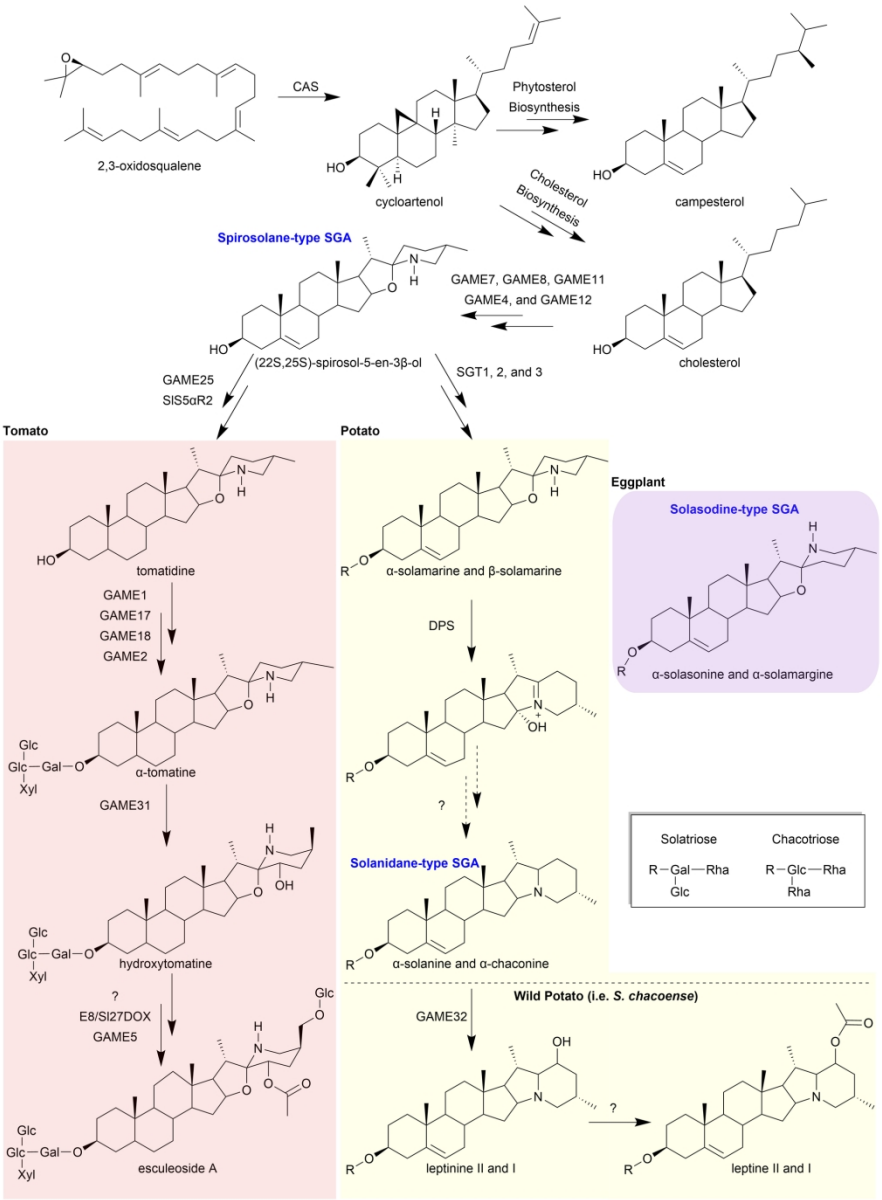


Figure 6. Steroidal glycoalkaloid biosynthesis in *Solanum*. CAS cyclizes 2,3-oxidosqualene from the mevalonate pathway to form cycloartenol a common metabolite in both phytosterol and cholesterol biosynthesis. Cycloartenol is converted to campesterol by a ten-step pathway and through a nine-step pathway to form cholesterol[121]. Following the production of cholesterol, five GAME enzymes are required to produce the spirosolane-type SGA core[8]. In tomato (red shaded box), GAME25 catalyzes the first of four steps resulting in tomatidine formation via the reduction of the spirosolane-type SGA core[123, 124]. Subsequent sugar additions by GAME1, GAME17, GAME18, and GAME2 result in the formation of α-tomatine[8]. GAME31, E8/Sl27DOX, GAME5, and an unknown acetyltransferase catalyze the fruit ripening associated formation of esculeoside A from α-tomatine[117, 126-129]. In potato (yellow shading), the addition of solatriose and chacotriose moieties by sequential sugar additions to (22*S*, 25*S*)-spirosol-5-en-3β-ol results in the formation of α- and β-solamarine, respectively[10]. The oxidization of α- and β-solamarine by DPS represents the first step in α-solanine and α-chaconine, Solanidane-type SGA, formation[10]. In *S.*

chacoense, α-solanine and α-chaconine are oxidized by GAME32 to form leptinines, and leptine formation requires the acetylation at the GAME32 introduced oxidation[117]. The solasodine-type SGAs (α-solasonine and α-solamargine) are the main SGAs in eggplant (purple shading) and contain solatriose and chacotriose moieties at the C-3 position, respectively. The biosynthetic mechanism leading to the stereochemical difference in spirosolane and solasodine cores remains uncharacterized[10, 120]. Enzyme abbreviations are as follows: CAS, cycloartenol synthase; GAME, glycoalkaloid metabolism; SlS5αR2, steroid 5α-reductase 2; SGT, solanidine glycosyltransferase; DPS, dioxygenase for potato solanidane synthesis; E8/Sl27DOX, α-tomatine 27-hydroxylase; Gal, galactose; Glc, glucose; Xyl, xylose; Rha, Rhamnose.

172x233mm (300 x 300 DPI)

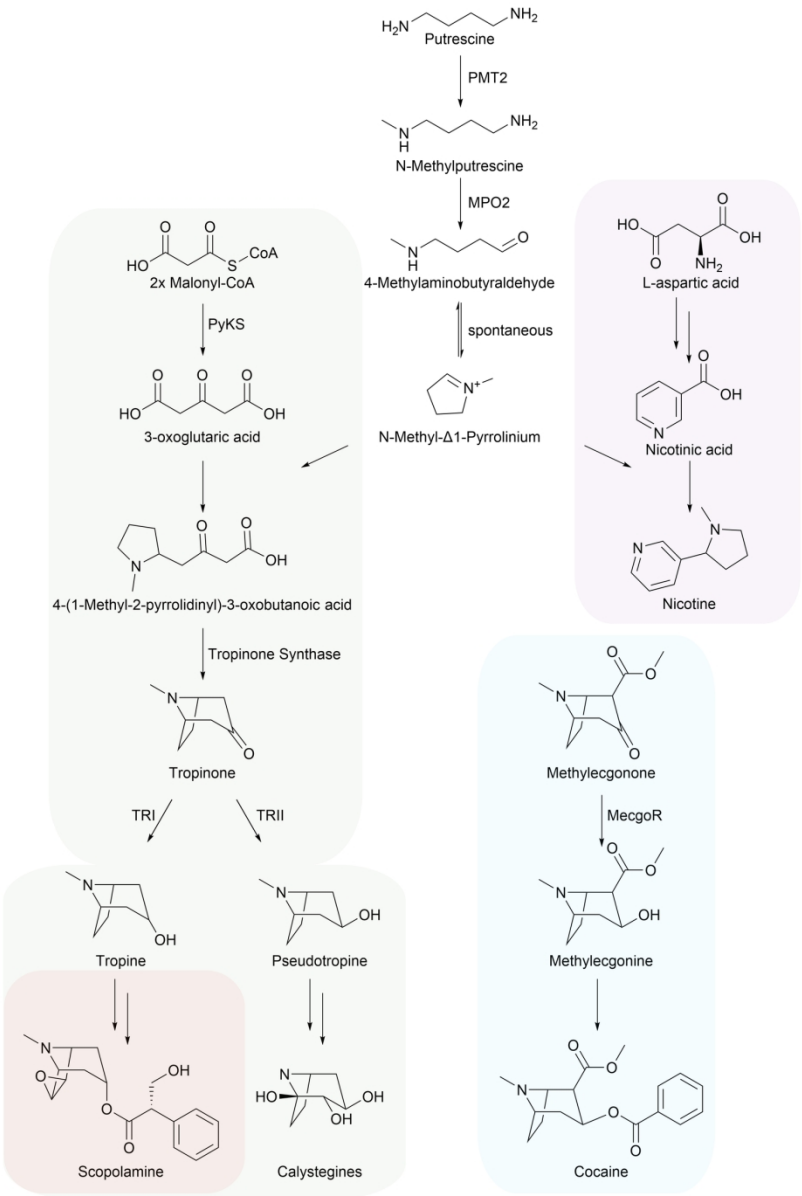


Figure 7. Evolutionary trajectories of tropane and nicotine formation in distinct plant lineages. Comparison of tropane and nicotine alkaloid biosynthesis reveals examples of both convergent (cocaine biosynthesis in *E. coca*) and divergent (nicotine biosynthesis) evolution[132, 142]. Scopolamine (orange) and nicotine (purple) represent alternative fates of the *N*-methylpyrrolinium cation in different genera of the Solanaceae. The use of an aldo-keto reductase enzyme (MecgoR) in the penultimate step of cocaine biosynthesis (blue) contrasts with catalysis by short-chain dehydrogenase/reductase (SDR) family enzymes (TRI and TRII) in scopolamine formation (green)[132]. *Not shown is catalysis by a single, bifunctional SDR to produce both tropine and pseudotropine in Brassicaceae[131]. Tropanol biosynthesis (green) is widely distributed across the Solanaceae compared to the biosynthesis of tropane aromatic esters such as scopolamine (orange)[137]. Enzyme abbreviations are as follows: PMT2, Putrescine *N*-methyltransferase 2; MPO2, *N*-methylputrescine oxidase 2; PyKS, Polyketide Synthase; TRI, Tropinone reductase I; TRII, Tropinone Reductase II; MecgoR, Methylecgonone reductase.

134x199mm (300 x 300 DPI)

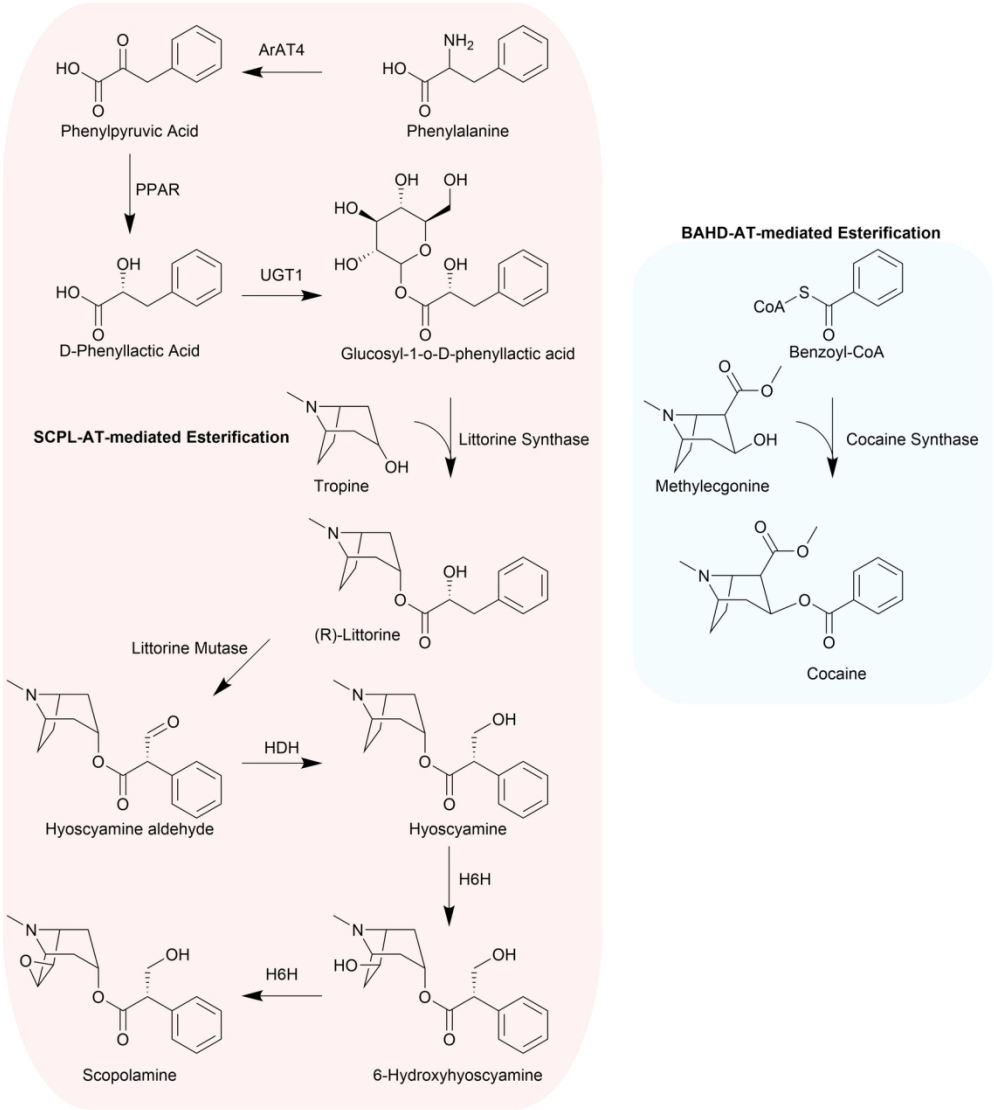


Figure 8. Independent evolution of tropane aromatic ester formation in Solanaceae and Erythroxylaceae. Scopolamine biosynthesis requires the biosynthesis of D-phenyllactic acid via a two-step process mediated by ArAT4 and PPAR[147, 148]. D-phenyllactic acid is glycosylated by UGT1 to form a glucose ester of phenyllactic acid, which is used, along with tropine, as substrate for littorine biosynthesis by Littorine Synthase, a serine carboxypeptidase-like acyltransferase[136]. Three enzymes, Littorine Mutase, HDH, and H6H, are required for the conversion of littorine to scopolamine[138, 150, 151]. In contrast, cocaine biosynthesis utilizes a BAHD acyl-transferase and coenzyme A donor to facilitate the transfer of a benzoyl moiety on to methylecgonine, the *E. coca* tropanol, to form cocaine[153]. Enzyme abbreviations are as follows: ArAT4, Aromatic amino acid transferase 4; PPAR, phenylpyruvic acid reductase; UGT1, UDP-glycosyltransferase 1; HDH, Hyoscyamine dehydrogenase; H6H, hyoscyamine-6-hydroxylase.

151x168mm (300 x 300 DPI)