Natural Product Reports

ROYAL SOCIETY OF CHEMISTRY

REVIEW

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Nat. Prod. Rep.*, 2022, **39**, 1438

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 [ab] and Cornelius S. Barry *[c]

Covering: 2000–2022

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.

Received 17th January 2022

DOI: 10.1039/d2np00003b

rsc.li/npr

[a]*Department of Biochemistry & Molecular Biology, Michigan State University, East Lansing, MI 48824, USA*
[b]*Department of Plant Biology, Michigan State University, East Lansing, MI 48824, USA*
[c]*Department of Horticulture, Michigan State University, East Lansing, MI 48824, USA. E-mail: barrycs@msu.edu*
† These authors contributed equally to this review.

Published on 24 March 2022. Downloaded by Michigan State University on 10/17/2023 4:02:57 PM.

This journal is © The Royal Society of Chemistry 2022

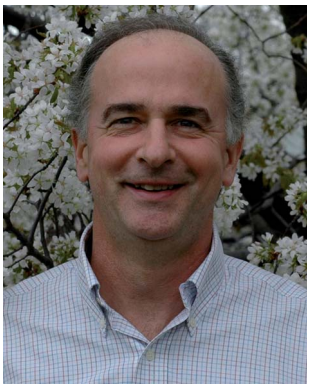
Published on 24 March 2022. Downloaded by Michigan State University on 10/17/2023 4:02:57 PM.

1 The Solanaceae: a phylogenetic framework for exploring metabolism

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

Although plant specialized metabolites exhibit considerable chemical complexity, they are ultimately derived from a pool of general metabolites formed through photosynthesis, glycolysis, the TCA cycle, amino acid metabolism, and the MEP-pathway.[5]

Paul Fiesel is a PhD candidate in the Department of Biochemistry and Molecular Biology at Michigan State University (East Lansing, MI, USA). In 2016, he received his Bachelor's degree (Biochemistry and Molecular Biology; Plant Biology) at the University of Minnesota. His research focuses on the evolution of acylsugars within the Solanum genus.



Robert Last is a University Distinguished Professor and Barnett Rosenberg Professor in the Department of Biochemistry and Molecular Biology and Department of Plant Biology at Michigan State University. He received his PhD in Biological Sciences from Carnegie Mellon in Pittsburgh, US, in 1986. He performed postdoctoral research at Whitehead Institute at MIT. Prior to joining the faculty at Michigan State University in 2004, he led research groups at the Boyce Thompson Institute at Cornell University and Cereon Genomics LLC. His current research focuses on the Nightshade Family as a model for understanding the evolution of metabolic form and function.

Hannah Parks is a PhD candidate in the Department of Biochemistry and Molecular Biology at Michigan State University. She received her BSc in biology at the University of Central Arkansas in 2018. Her research focuses on metabolite discovery and biosynthetic pathway elucidation in Atropa belladonna.

Cornelius Barry is an Associate Professor in the Department of Horticulture at Michigan State University, USA. He received his PhD in 1995 from the University of Nottingham, UK, studying ethylene biosynthesis in tomato and performed postdoctoral research investigating the genetic mechanisms underlying fruit ripening at the University of Nottingham, Texas A&M University, USA, and the Boyce Thompson Institute for Plant Research, USA. He joined the faculty at Michigan State University in 2007. His current research explores the biochemical evolution of specialized metabolism in the Nightshade family, with a focus on tropane alkaloids.

Published on 24 March 2022. Downloaded by Michigan State University on 10/17/2023 4:02:57 PM.

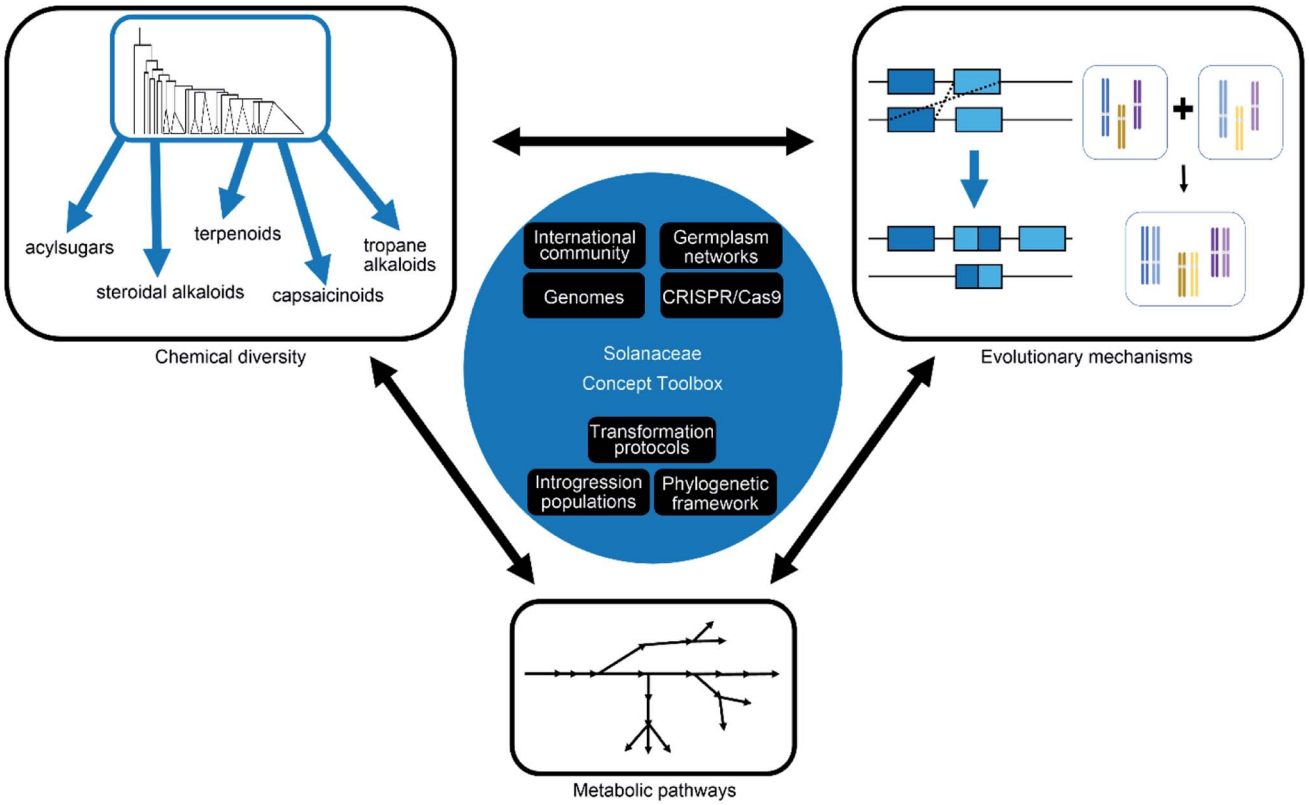


Fig. 1 Solanaceae as a model family for specialized metabolism evolution studies. The Solanaceae concept toolbox connects biodiversity, genetics, and evolutionary mechanisms to each other. Chemical diversity informs metabolic pathway discovery, which in turn reveals evolutionary mechanisms underlying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

General metabolites undergo transformations, including ligation and cyclization to generate scaffold molecules that are modified by glycosylation, acylation, methylation, prenylation, oxidation, and reduction to dramatically increase chemical complexity. In 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 (https://www.solanaceaesource.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, pangenomes, 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.

 This journal is © The Royal Society of Chemistry 2022

Published on 24 March 2022. Downloaded by Michigan State University on 10/17/2023 4:02:57 PM.

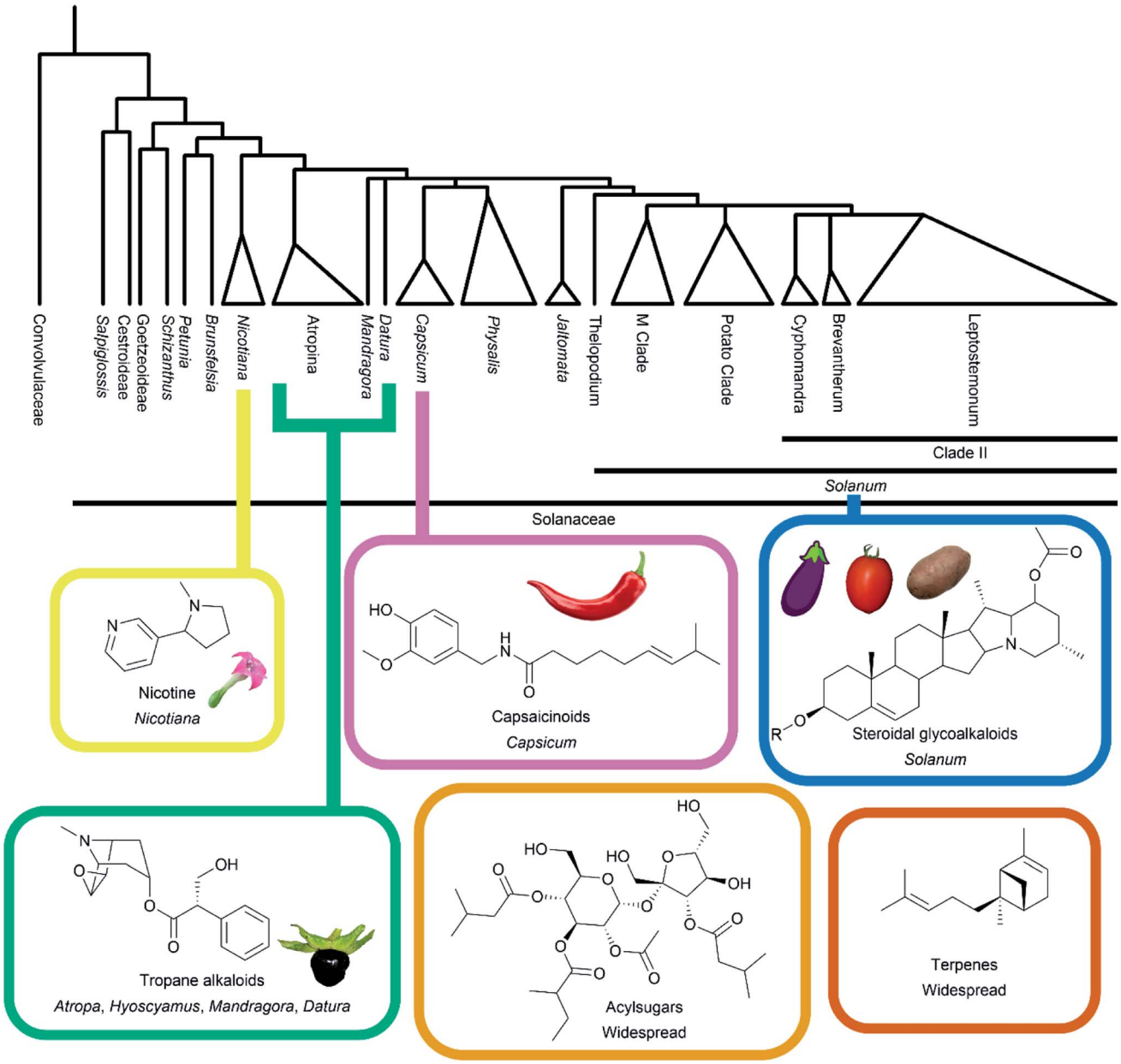


Fig. 2 Phylogenetic distribution of major Solanaceae specialized metabolite classes. The Solanaceae family produces specialized metabolites of multiple chemical classes. A simplified phylogeny of the Solanaceae family is shown based on prior determination of phylogenetic relationships.[11,12] Major metabolite 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.

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

The Solanaceae has emerged as a model system for investigating the biosynthesis and evolution of specialized

Published on 24 March 2022. Downloaded by Michigan State University on 10/17/2023 4:02:57 PM.

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

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

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

Recent progress in understanding the genetic and biochemical basis of tomato flavor was facilitated by large-scale genome sequencing and resequencing projects involving hundreds of phenotypically diverse cultivated tomato accessions and wild relatives. These studies revealed insights into the nature of the tomato pan-genome and sequence variation associated with crop domestication and improvement, including gene duplication, single nucleotide polymorphisms, insertion–deletions, and large-scale structural variants.[16,17,30,31] The development of these resources facilitates the identification of genetic variation underlying phenotypic traits *via* genome-wide association studies. Notably, this approach was successfully deployed for the identification of genetic components underlying variation in tomato fruit flavor and aroma, revealing how human selection for visible traits such as fruit size, yield, and color can lead to alternative outcomes and unintentionally influence SM pathways that contribute to fruit quality.

Several hundred volatiles are detectable in ripening tomato fruits, but consumer taste panels identified 33 metabolites associated with consumer liking and 37 correlated with flavor intensity.[30] These influential aroma volatiles are derived through diversion of general metabolites, including carotenoids, phenylalanine, isoleucine/leucine, and fatty acids into diverse SM pathways. Genetic variation is evident across tomato varieties and 13 fruit aroma volatiles are significantly reduced in a collection of 48 modern cultivars when compared to 236 heirloom tomato varieties. This work shows that breeding of modern varieties for traits such as yield, shelf-life, and disease resistance has inadvertently and negatively altered SM pathways that produce aroma volatiles associated with consumer preference.[30] Subsequent GWAS analyses performed using a panel of 398 diverse tomato accessions analyzed for 27 volatiles along with glucose, fructose, malic acid, and citric acid revealed the existence of 251 association signals for 20 traits, including 15 correlated with aroma volatile production.

Among these associations are five loci that influence the production of carotenoid-derived volatiles. Two loci specifically influence the production of geranylacetone, which is formed by oxidative cleavage of the minor tomato fruit carotenoids phytoene, phytofluene, ζ-carotene, and neurosporene. A single locus specifically influences 6-methyl-5-hepten-2one (MHO) accumulation, which is derived from lycopene, the main carotenoid pigment in red-fruited tomato varieties. Two additional loci are associated with the production of both geranylacetone and MHO. Analysis of allele frequencies at these loci indicate that genetic complexity was progressively lost during breeding to the point where essentially only two allele combinations associated with accumulation of both volatiles persist in most modern cultivars. Analysis of MHO levels in genotypes with distinct allele combinations revealed that, as breeders selected for high lycopene in red-fruited varieties, they inadvertently selected favorable alleles that increase MHO production. In contrast, the favorable alleles that promote geranylacetone accumulation are absent in modern cultivars.[30]

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

The phenylalanine-derived volatiles guaiacol, eugenol, and methylsalicylate contribute to the aroma of tomato fruits and are associated with smoky and medicinal-like aromas, which are often negatively correlated with consumer liking.[34] Guaiacol, eugenol, and methylsalicylate accumulate in tomato fruits as

 This journal is © The Royal Society of Chemistry 2022

Published on 24 March 2022. Downloaded by Michigan State University on 10/17/2023 4:02:57 PM.

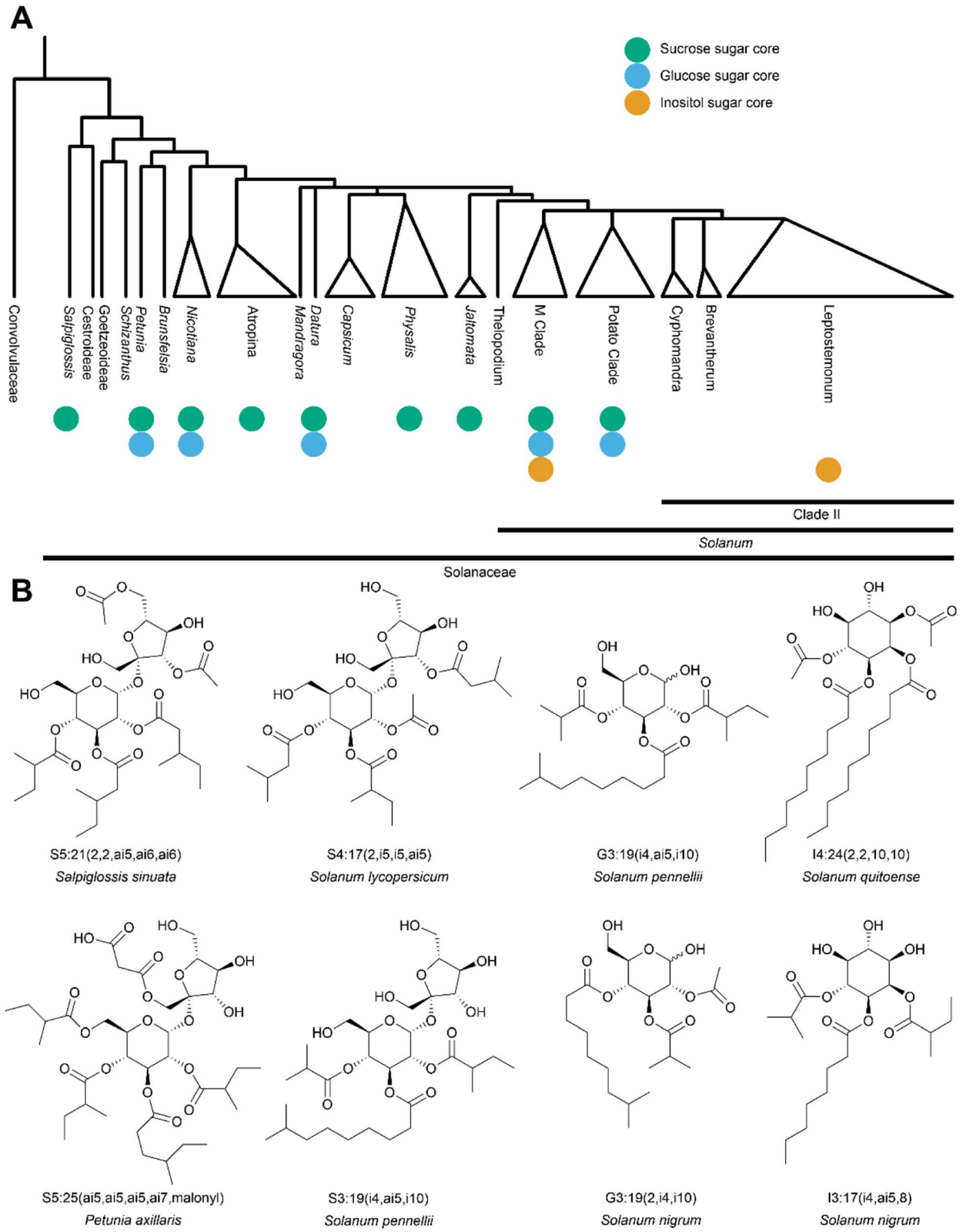


Fig. 3 Phylogenetic distribution of acylsugar core types. (A) Simplified Solanaceae phylogeny with acylsugar core type placed on each lineage with characterized acylsugars. The phylogenetic tree is based upon previously published Solanaceae and *Solanum* trees.[11,12]. (B) Characteristic acylsugar structures 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).

Published on 24 March 2022. Downloaded by Michigan State University on 10/17/2023 4:02:57 PM.

diglycosides, and cleavage of the glycoside groups leads to release of the volatiles in "smoky" cultivars. In contrast, in "non-smoky" varieties these metabolites exist as non-cleavable triglycosides resulting in reduced levels of volatile release.[35] Formation of guaiacol, eugenol, and methylsalicylate triglycosides from their diglycoside precursors is catalyzed by the UDP-glucosyltransferase enzyme, NON-SMOKY GLYCOSYLTRANSFERASE1 (NSGT1). The *NSGT1* gene resides at a locus on chromosome 9 that contains a second gene designated *NSGT2*. Both genes contain structural changes in "smoky" cultivars that are predicted to render them non-functional although the exact structure of the locus was unresolved.[35]

The recent development of 14 new reference tomato genomes assembled using Oxford Nanopore long read sequencing technology allowed the genome structure flanking the *NSGT1* locus to be resolved. Five haplotypes were identified revealing evidence of intraspecific gene duplication and loss at an SM locus that was selected during crop improvement.[17] Haplotype I is proposed to be ancestral and contains predicted functional copies of *NSGT1* and *NSGT2*. All other haplotypes contain coding sequence mutations in *NSGT2*. In addition, haplotypes IV and V also lack functional copies of *NSGT1* and are therefore null mutations for both *NSGT1* and *NSGT2*. Analysis of guaiacol levels across two GWAS panels and within an F_2 population segregating for haplotype V and a functional copy of *NSGT1* demonstrated that fruit guaiacol levels are reduced in individuals that contain a functional copy of *NSGT1*. Together, these data illustrate the combined power of genome sequences developed using long-read sequencing data and GWAS to investigate the evolution of loci associated with SM phenotypes, particularly when the variation is mediated by tandem gene duplication that may be unresolved in genome assemblies derived from short-read data. Overall, these studies represent an example of fundamental science that provides opportunities to breed tomato varieties with favorable aroma volatile alleles.

3 Sticky: single-cell biochemical genetics reveals acylsugar metabolic complexity

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

3.1 Harnessing acylsugar genotypic diversity for tomato pathway determination

Tomato acylsugar diversity was employed to uncover the acylsugar biosynthesis pathway within cultivated tomato, *S. lycopersicum*. Analysis of *S. lycopersicum* introgression lines carrying *S. pennellii* chromosomal segments was instrumental in identifying loci required for acylsugar biosynthesis.[24,62] The identification and subsequent validation of candidate genes was facilitated by trichome-specific transcriptome, *in vitro* enzyme assays, and *in vivo* gene VIGS knockdown and CRISPR/Cas9 knockout. These approaches uncovered the core acylsugar pathway in *S. lycopersicum* glandular trichomes. A series of evolutionarily related BAHD acyltransferases, named AcylSucrose AcylTransferase 1–4 (ASAT1-4), acylate sucrose sequentially to produce tetraacylsucroses consisting of acyl chains at R_2, R_3, R_4, and R_3,[24,63,64] (Fig. 4). Each enzyme selectively acylates specific sucrose hydroxyls with varying promiscuity for acyl-CoA substrates. Documenting this pathway enabled discovery of mechanisms responsible for acylsugar diversity in wild tomato relatives.

Intra- and inter-specific differences in tomato acylsugar structures result in part from differing ASAT activities. Comparative biochemical analysis of cultivated and wild tomato ASAT sequences uncovered amino acid residues responsible for specific activity differences. For example, the comparison of ASAT2 sequences and *in vitro* enzyme activities across tomato species revealed two mutations that impact acyl-CoA specificity. Residues Val/Phe[408] and Ile/Leu[44] influence the ability to use the structurally similar iC5-CoA and aiC5-CoA, respectively, without altering activity with nC12-CoA.[64] Comparison of *S. lycopersicum* and *S. habrochaites* ASAT3 homologs revealed a Tyr/Cys[41] residue change impacting the enzyme's ability to use nC12-CoA.[63] Characterization of *S. habrochaites* ASAT4 in accessions collected from Ecuador to Southern Peru revealed variations in acetylation patterns that were explained either by changes in ASAT4 expression or coding sequence mutations.[65,66] The comparative biochemistry approach revealed differences in enzyme acyl donor specificity, which impacted acylsugar phenotypes. This approach also determined evolutionary changes in enzyme acyl acceptor specificity.

S. pennellii LA0716 produces acylsucroses through a 'flipped pathway', resulting from changes in ASAT acyl acceptor specificity.[67] While cultivated tomato produces acylsucroses with one furanose ring acylation (termed F-type acylsucroses), *S. pennellii* and some *S. habrochaites* accessions synthesize acylsucroses acylated exclusively on the pyranose ring.[63] These 'P-type' acylsucroses are produced by alternate ASAT2 and ASAT3 homologs, which catalyze the third and second pathway steps, respectively. The published results suggest that *S. pennellii* ASAT2 likely evolved from an ancestral enzyme capable of acylating both mono- and diacylsucrose. Analogous sequence

 This journal is © The Royal Society of Chemistry 2022

Published on 24 March 2022. Downloaded by Michigan State University on 10/17/2023 4:02:57 PM.

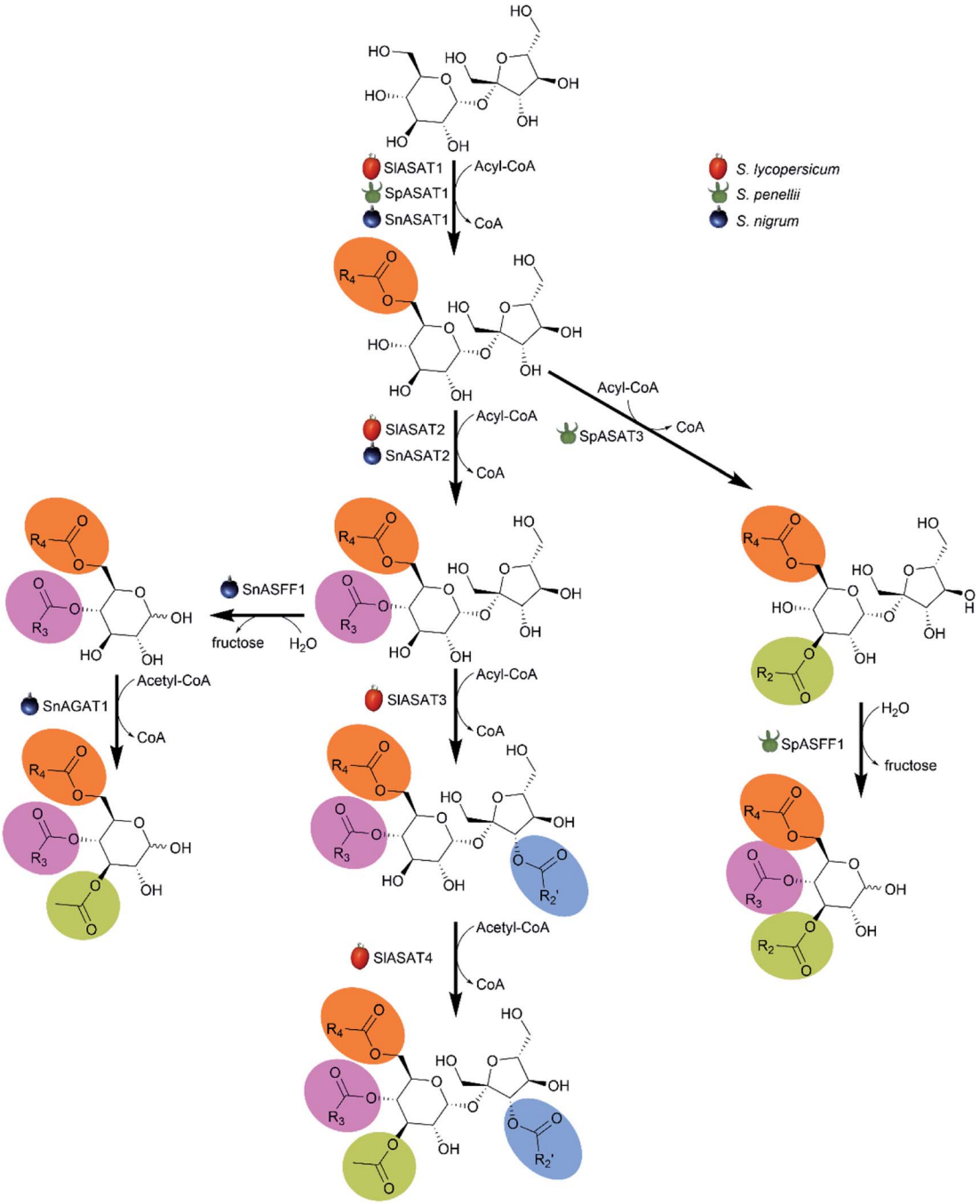


Fig. 4 Acylsucrose and acylglucose pathway diversity in *Solanum* species. The acylsucrose and acylglucose biosynthesis pathways for *S. nigrum*, *S. lycopersicum* and *S. pennellii*. All three biosynthetic pathways begin by acylating sucrose.[24,63,64,68,72] Sequential acylations produce tetraacylsucroses, 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.

changes in ASAT3, potentiated by ASAT3 duplication, resulted in the neofunctionalized ASAT3 duplicate found in *S. habrochaites* and *S. pennellii*. This study revealed a remarkably small number of amino acid changes that caused a major change in pathway structure and product phenotypes in closely related species.

The flipped *S. pennellii* pathway and recruitment of an invertase-like enzyme appear to have potentiated evolution of *S. pennellii* acylglucose synthesis (Fig. 4). *S. pennellii* acylglucoses are synthesized from P-type acylsucroses by a neofunctionalized glycoside hydrolase 32 family (GH32) beta-fructofuranosidase, SpASFF1.[68] The modified SpASFF1

Published on 24 March 2022. Downloaded by Michigan State University on 10/17/2023 4:02:57 PM.

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

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

Further evidence for this hypothesis was provided by identification of natural chain diversity associated with allelic diversity of two acyl-CoA biosynthesis genes.[70] These trichome-expressed genes, an enoyl-CoA hydratase (AECH1) and acyl-CoA synthetase (AACS1), reside in a gene cluster syntenic to the chromosomal region containing ASAT1. The Solanaceae family shares the syntenic region, which was likely derived from a Solanaceae-specific polyploidy event. Silencing AECH1 and AACS1 in *S. lycopersicum*, *S. pennellii*, and the more distantly related *Solanum quitoense*, reduced or eliminated medium length (10–12 carbons) acyl chains from acylsugars. Additionally, the presence of AECH1 and AACS1 correlates with natural variation in medium acyl chains. For example, in the short chain producing genera *Petunia* and *Nicotiana*, AECH1 and AACS1 are either missing or present as pseudogenes. These genes represent another example of how evolutionary changes in metabolic machinery impacted acylsugar composition.

3.2 Genomics tools enable comparative biochemistry in non-model organisms

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

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

The ancestral and derived acylsucrose pathways provide insight into the evolutionary origins of acylsugars.[37] Lamiidae BAHD sequence homology, phylogenetics, and known whole genome duplication events all enabled inferences regarding

 This journal is © The Royal Society of Chemistry 2022

Published on 24 March 2022. Downloaded by Michigan State University on 10/17/2023 4:02:57 PM.

early acylsugar evolution. One hypothesis, based on sequence analysis, is that ASAT sequences derive from an alkaloid biosynthetic enzyme ancestor. Based on nonsynonymous mutation rates and historical polyploidy events, the clade containing ASAT1,2,3 appears to have arisen *via* an ancient whole genome duplication before the Solanaceae-Convolvulaceae split (~50–65 MYA). Subsequent duplications prior to, and following the Solanaceae polyploidization, led to evolution of the ASATs and paralogs found in the ASAT1,2,3 clade. As described above, our model of acylsugar biosynthetic pathway evolution invokes loss of aASAT1, refinement of ASAT1 and ASAT2 activities, and recruitment of ASAT3 occurred later in Solanaceae diversification.

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

In contrast to the detailed information available for acylsucrose and acylglucose biosynthesis, the pathway leading to acylinositol synthesis in the *Solanum* remains largely enigmatic. So far only one enzyme was demonstrated in acylinositol biosynthesis: the *S. quitoense* enzyme TriAcylInositol AcetylTransferase, SqTAIAT, acetylates triacylinositols to produce tetraacylinositols.[71] SqTAIAT is the closest known *S. quitoense* homolog to the final enzyme in tomato acylsucrose biosynthesis, SlASAT4, indicating conservation of acetyltransferases across acylinositol and acylsucrose biosynthesis. Both enzymes acetylate triacylsugars differing in their sugar core. Similar enzymatic activity and high sequence similarity suggest a common evolutionary origin for acylinositol and acylsucrose biosynthesis. However, the initial steps of acylinositol biosynthesis remain unresolved. Further pathway elucidation in *S. quitoense* and *S. nigrum* may uncover the evolutionary innovations underlying acylinositol production.

3.3 Into the depths with acylsugars

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

4 Stinky: variations on a theme define terpene diversity across *Solanum*

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

Plant terpenoid diversity is created at multiple levels. Firstly, small gene families produce *cis* and *trans*-prenyltransferases that initially condense a single molecule of DMAPP and IPP to form either geranyl diphosphate (GPP) (*trans* isomer) or neryl diphosphate (NPP) (*cis* isomer). These C10 metabolites can then be extended by five carbon units, through condensation with additional units of IPP, to yield *trans*- or *cis*-farnesyl diphosphate (*E*,*E*-FPP or *Z*,*Z*-FPP, C15), geranylgeranyl or nerylneryl diphosphate (GGPP or NNPP, C20), or longer chain prenyl diphosphates.[79] Short-chain prenyl diphosphates (C10–C20) are substrates for terpene synthases (TPS), which exist as moderately large gene families (up to ~100 members) and catalyze the formation of hydrocarbon terpene skeletons *via* rearrangements and cyclization. TPS enzymes possess considerable catalytic potential. They frequently utilize more than one substrate, and catalysis by a single enzyme often generates multiple products.[79–81] These hydrocarbon terpene skeletons are often functionalized by the addition of hydroxyl groups, which provide targets for modifications such as epoxidation,

Published on 24 March 2022. Downloaded by Michigan State University on 10/17/2023 4:02:57 PM.

methylation, acylation, and glycosylation, ultimately generating the vast complexity of terpenoids observed across the plant kingdom.

The availability of a high-quality reference genome assembly for cultivated tomato (*Solanum lycopersicum*) facilitated what is likely the most comprehensive published catalogue of terpene scaffold biosynthesis in plants. The data highlight considerable chemical complexity with *in vitro* biochemical data revealing the potential to synthesize 53 known hydrocarbon terpene scaffolds plus several unidentified products. These terpenes arise through combined catalysis of seven *cis*-prenyltransferases and 10 *trans*-prenyltransferases that form C10, C15, and C20 prenyl diphosphates, together with 34 functional TPS enzymes.[82,83] Consistent with the known catalytic promiscuity of TPS enzymes, many of the tomato TPSs can utilize more than one substrate, particularly the sesquiterpene synthases that use both *E*,*E*-FPP and *Z*,*Z*-FPP, and yield multiple products. In addition, considerable catalytic redundancy exists. For example, eight distinct TPSs catalyze the formation of the monoterpene β-myrcene. Individual CPT, TPT, and TPS enzymes are localized to the cytosol, plastids, as well as mitochondria, and the corresponding genes are differentially expressed across tomato tissues: this highlights the spatial separation of terpene synthesis modules across tomato. Metabolite profiling of 13 tomato tissues identified 29 out of 53 terpenes *in planta*, suggesting that some terpenes are either below the limit of detection in tomato grown under standard cultural conditions or are further modified to produce more structurally complex metabolites.

Genomic clustering is a key feature of terpene biosynthetic genes in plants.[84] These clusters generally consist of both paralogs and non-homologous genes encoding enzymes of terpene biosynthesis, creating a reservoir for the evolution of chemical novelty and facilitating the inheritance of SM modules that promote plant adaptation. Gene duplication within these clusters is often followed by pseudogenization and gene loss to create additional chemical variation. The majority of the 52 TPS loci in tomato, including 18 predicted pseudogenes, are located within gene clusters dispersed across the genome.[82] In addition, the *TPS* gene clusters on chromosomes 6, 7, 8, and 12 also contain combinations of *cis* or *trans* prenyltransferases, cytochromes P450, methyltransferases, acyltransferases, and glycosyltransferases.[82,85] While most of the potential terpene modifying enzymes within these clusters await functional characterization, a three-gene subcluster on chromosome 8 comprising *SlTPS21-CYP71D51-SlCPT2* was demonstrated to synthesize (+)-lycosantalonol from NNPP.[86]

Along with the existence of the 18 *TPS* pseudogenes in the tomato genome, three *TPS*-related gene clusters on chromosomes 6, 8, and 12 also contain inactive cytochromes P450 genes.[82] The high prevalence of pseudogenes within these tomato terpene biosynthetic gene clusters suggests that there is potential for considerable genetic variation. For example, a gene that is pseudogenized in one accession or species may be functional in another. Thus, variation in terpene-related gene clusters may exist between distinct accessions of *S. lycopersicum* but also more likely across the genomes of diverse Solanaceae species. The increasing availability of high-quality chromosome scale reference genomes assembled from long-read sequencing will facilitate identification of additional gene clusters and future comparative evolutionary analysis of terpene biosynthesis across the Solanaceae.

Within the *Solanum* genus, distinct evolutionary trajectories associated with trichome-derived terpene-related gene clusters are indeed apparent between cultivated tomato and wild relatives that diverged from a common ancestor approximately two-three million years ago.[11] Notably, while limited terpene diversity exists in trichomes between cultivated tomato accessions, considerable variation is observed across distinct populations of *Solanum habrochaites* and between *S. habrochaites* and *S. lycopersicum*.[87] This genetic variation determines whether specific accessions preferentially synthesize monoterpenes (C10) or sesquiterpenes (C15), and results from differences at the *cis-prenyltransferase 1* (*CPT1*) locus and associated TPS-e/f enzymes that are located within the chromosome 8 terpene gene cluster.[85] For example, trichomes of cultivated tomato predominantly accumulate the monoterpene β-phellandrene, which is synthesized from NPP by neryl diphosphate synthase1 (NDPS1).[88] While select monoterpene-producing accessions of *S. habrochaites* also contain an ortholog of NDPS1, a separate group of sesquiterpene producing accessions of *S. habrochaites* possess the C15-producing *Z*,*Z*-farnesyl diphosphate synthase (zFPS) at the *CPT1* locus[89,90] (Fig. 5). Comparative sequence analysis, homology modeling, and site-directed mutagenesis revealed that the relative positioning of bulky aromatic amino acid residues within a hydrophobic cleft specifies substrate binding and prenyl-chain elongation between CPT1 isoforms with NDPS1 and zFPS activity and that this contributes to intraspecific terpene variation in *S. habrochaites*.[90]

Together with divergent CPT1 enzymes, terpene diversity in *S. habrochaites* trichomes is also driven by natural variation in chromosome 8 cluster TPS-e/f subfamily members. *S. lycopersicum*, synthesizes a cocktail of monoterpenes in trichomes from NPP using the TPS-e/f enzyme, β-phellandrene synthase (SlPHS1/SlTPS20).[88] PHS1 activity is conserved in some *S. habrochaites* accessions while others contain the TPS-e/f paralogs limonene synthase (ShLMS) and pinene synthase (ShPIS), which catalyze the formation of limonene and α-pinene from NPP, respectively.[87] In addition to this intraspecific variation in monoterpene biosynthesis, two additional groups of *S. habrochaites* accessions possess TPS-e/f enzymes that synthesize sesquiterpenes from *Z*,*Z*-FPP produced by zFPS: santalene and bergamotene synthase (ShSBS) catalyzes the formation of a mixture of santalene and bergamotene isomers.[87,89] In contrast, a distinct, yet closely related enzyme, zingiberene synthase (ShZIS) catalyzes the formation of 7-epizingiberene[87] (Fig. 5). These sesquiterpene forming TPS-e/f enzymes are not present in *S. lycopersicum* and, to date, appear to be restricted to a subset of *S. habrochaites* accessions. Overall, together with variation at the *CPT1* locus, these examples illustrate the evolutionary potential of SM associated gene clusters to create and maintain inter-specific and intra-specific chemical diversity. This relatively rapid intra-specific evolution of chemical

 This journal is © The Royal Society of Chemistry 2022

Published on 24 March 2022. Downloaded by Michigan State University on 10/17/2023 4:02:57 PM.

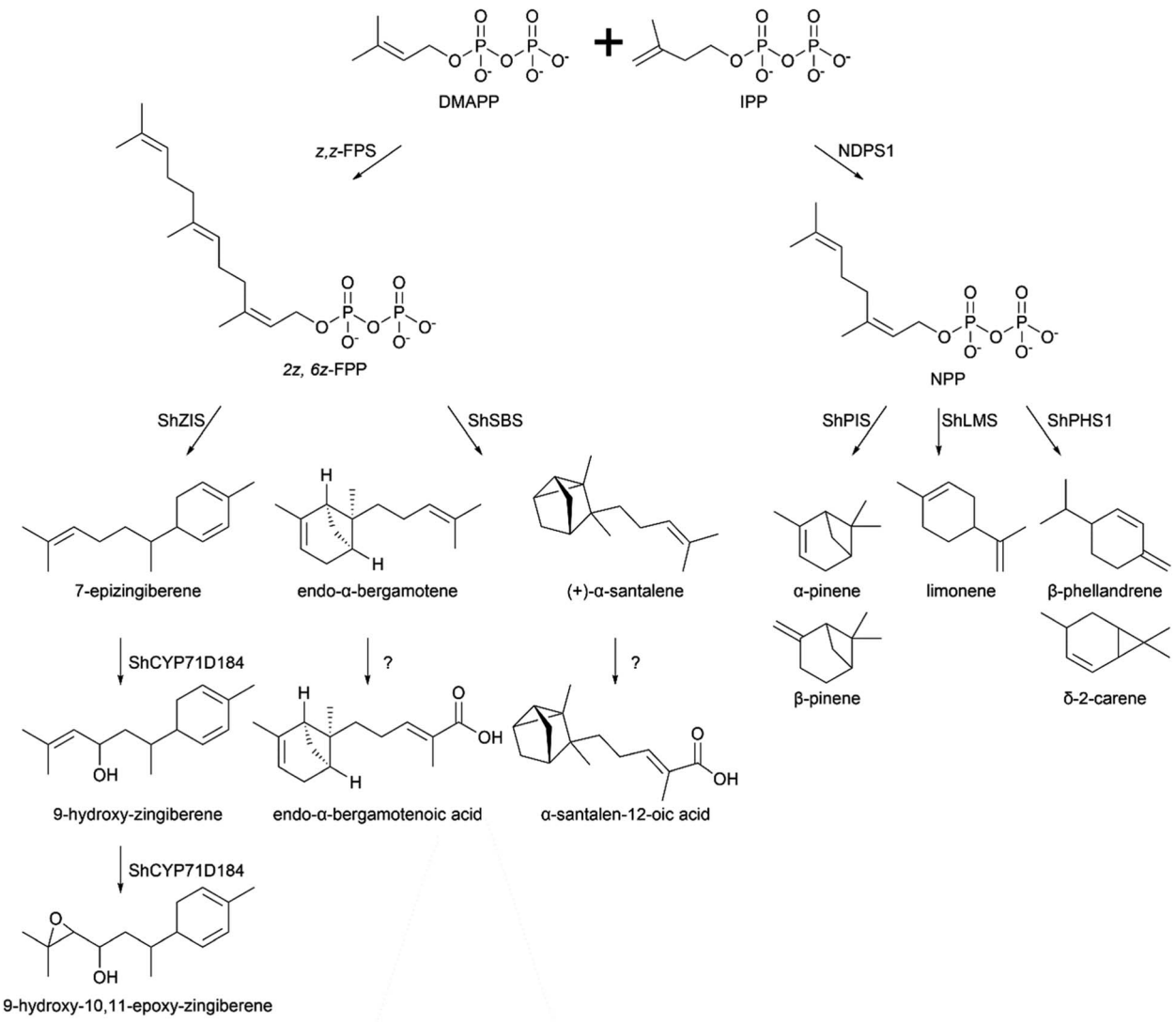


Fig. 5 Terpenoid biosynthesis in the trichomes of *Solanum habrochaites* derived from *cisoid* substrates. NDPS1 catalyzes the condensation of a single molecule of DMAPP and IPP to form NPP (C10).[88] In contrast, *z,z*-FPS catalyzes the formation of *2z,6z*-FPP (C15) through sequential condensation 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.

variation in specific populations of plants may confer selective advantage against diverse biotic challenges.

The ability of trichomes of select *S. habrochaites* accessions to synthesize the sesquiterpenes santalene and bergamotene as well as 7-epizingiberene and their derivatives is known to confer increased tolerance to insect pests and pathogens when compared to trichomes that synthesize *S. lycopersicum* type monoterpenes.[91–94] Santalene and bergamotene backbones are oxidized into sesquiterpene acids *via* unknown enzymes.[93] In contrast, 7-epizingiberene is sequentially oxidized to a combination of 9-hydroxy-zingiberene and 9-hydroxy-10,11-epoxy-zingiberene in reactions catalyzed by the trichome-expressed cytochrome P450, ShCYP71D184 (ref. 95) (Fig. 5). 9-Hydroxy-10,11-epoxy-zingiberene is particularly effective in bioactivity assays against whiteflies (*Bemisia tabaci*) and the microbial pathogens, *Phytophthora infestans* and *Botrytis cinerea*. ShCYP71D184 is encoded by the *Sohab01g008670* locus and is therefore not located in the chromosome 8 TPS cluster responsible for the synthesis of the 7-epizingiberene substrate. The predicted ShCYP71D184 protein is 94% identical to its

Published on 24 March 2022. Downloaded by Michigan State University on 10/17/2023 4:02:57 PM.

putative ortholog from *S. lycopersicum* SlCYP71D184/Solyc01g008670. The function of SlCYP71D184 is unknown but *S. lycopersicum* trichomes do not synthesize 7-epi-zingiberene and this enzyme is incapable of catalyzing the formation of 9-hydroxy-zingiberene and 9-hydroxy-10,11-epoxy-zingiberene. Although not completely understood, these data suggest that, like other loci that influence terpene biosynthesis in glandular trichomes of *Solanum*, genetic variation exists at the *CYP71D184* locus that specifies chemical diversity.

5 Spicy: lineage-specific biosynthesis of capsaicinoids in pepper

Species within the *Capsicum* genus of the Solanaceae possess the capacity to synthesize a group of specialized metabolites known as capsaicinoids, including capsaicin, the principal determinant of pungency in chili peppers. These specialized metabolites are of culinary and cultural importance but also possess applications as topical pain medications and show efficacy as anti-inflammatories, treatments for cancer and weight-loss, and possess anti-microbial activities.[96–99] Capsaicinoids are synthesized within the placenta that surrounds the seeds of developing fruit and act as feeding deterrents for small mammals such as rodents, but not birds.[100] This deterrence is mediated by the mammalian vanilloid receptor 1 (VR1) ion channel that is localized to sensory nerve endings and responds to heat stimuli.[101] The ortholog of VR1 from birds does not respond to capsaicin and as such, birds, which are more efficient seed dispersers than small mammals, are unaffected by the pungency of pepper fruits.[102]

The biosynthesis of capsaicinoids is not fully understood, particularly at the biochemical level and this pathway is yet to be reconstructed in a heterologous system. However, capsaicin biosynthesis is considered a derived trait within *Capsicum*, as species from the more ancient Andean clade of the genus are non-pungent.[103] Within *Capsicum* species, intra-specific variation exists resulting in loss of pungency.[103] Most notably, this intra-specific variation occurs in the major crop species *Capsicum annuum* and gives rise to both pungent and sweet pepper cultivars.[103] Capsaicin is synthesized through the condensation of vanillylamine, derived from the phenylpropanoid pathway, with 8-methyl-6-nonenoyl-CoA, produced through branched-chain amino acid metabolism and fatty acid synthesis.[104] Genetic analyses identified loci associated with capsaicin accumulation and genes within the phenylpropanoid, branched-chain amino acid catabolism, and fatty acid synthesis pathways are among the candidates discovered.[105–107] For example, loss of function alleles at the *AMT* locus, which encodes an aminotransferase that catalyzes the formation of vanillylamine from vanillin, disrupts capsaicin biosynthesis.[108–110] Similarly, mutation in a ketoacyl-ACP reductase (*CaKR1*), an enzyme involved in fatty acid biosynthesis, resulted in undetectable levels of capsaicin and 8-methyl-6-nonenoic acid, a precursor of 8-methyl-6-nonenoyl-CoA.[111] In addition, the BAHD acyltransferase capsaicin synthase, also known as Pun1, is associated with pungency in hot pepper and proposed to catalyze the condensation of vanillylamine with 8-methyl-6-nonenoyl-CoA to form capsaicin.[112] A 2.5 kb deletion allele at this locus is present in non-pungent genotypes, although biochemical evidence supporting a direct role for this enzyme in capsaicin biosynthesis is lacking.[112] Overall, these studies reveal genetic variation across *Capsicum* that has likely arisen due to domestication and selection.

6 Bitter: evolutionary signatures of glycoalkaloid biosynthesis in *Solanum*

Steroidal glycoalkaloids (SGAs) are bitter and toxic metabolites that occur in *Solanum* including the crop species tomato, potato, and eggplant. SGAs provide protection against herbivory as well as microbial pathogens and are proposed to function through the disruption of cell membranes and inhibition of cholinesterase activity.[113] In the United States, SGA levels are monitored in potato to maintain levels below an FDA-regulated threshold due to their toxicity.[114] Evolution and domestication shaped SGA diversity in *Solanum*; metabolite profiling and chemical structure elucidation reveal hundreds of SGAs that differ among members of the genus due to gene gain and loss between species.[115,116] For example, α-tomatine and esculeoside A accumulate in tomato while α-solasonine and α-solamargine are synthesized in eggplant. In contrast, domesticated potato synthesizes α-solanine and α-chaconine, while leptines, SGAs that display efficacy against Colorado potato beetle (CPB), are found in wild potato species (Fig. 6).[10,117–120] SGAs arise from the modification of cholesterol produced from the mevalonate pathway and are characterized by a nitrogen-containing 27-carbon core, which can undergo multiple glycosylations to form steroidal glycoalkaloids.[121] Comparison of genomic sequences between species revealed that several biosynthetic steps of SGA formation in tomato, potato, and eggplant, encoded by *GLYCOALKALOID METABOLISM* (*GAME*) genes, are clustered within these genomes.[8,122]

Formation of plant SGA sterol cores requires diversion of 2,3-oxidosqualene from the mevalonate pathway into cholesterol biosynthesis, and this biosynthetic pathway appears to have evolved from the duplication and divergence of genes involved in phytosterol biosynthesis, which leads to the production of brassinosteroids, an essential class of phytohormones.[121] Cycloartenol synthase (CAS) converts 2,3-oxidosqualene into cycloartenol, and this metabolite is the branch point between cholesterol and phytosterol biosynthesis as it serves as a substrate for both SSR2 (sterol side chain reductase 2) and SMT1 (sterol C-24 methyltransferase) to form cycloartanol or 24-methylenecycloartanol, respectively.[121] Cholesterol biosynthesis leads to the production of the SGAs and saponins in both glycosylated and aglycone forms.[121] Elucidation of cholesterol biosynthesis in plants revealed five enzymes shared between the cholesterol and phytosterol pathways.[121] Phylogenetic analysis of enzymes specific to cholesterol biosynthesis suggests that C5-SD2 (sterol C-5(6) desaturase), 7-DR2 (7-dehydrocholesterol reductase), SMO3 (C-4 sterol methyl oxidase) and SMO4 likely

 This journal is © The Royal Society of Chemistry 2022

Published on 24 March 2022. Downloaded by Michigan State University on 10/17/2023 4:02:57 PM.

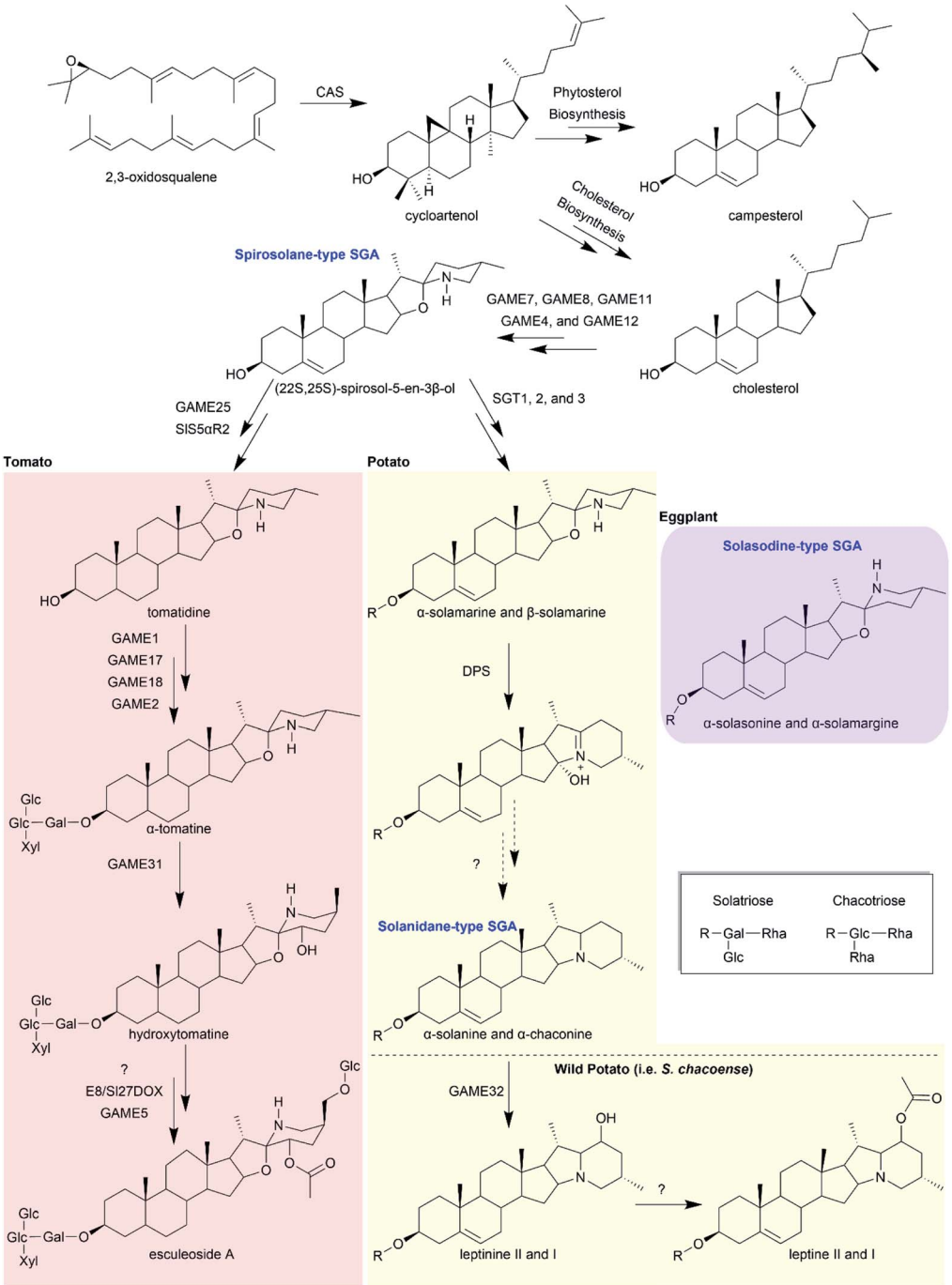


Fig. 6 Steroidal glycoalkaloid biosynthesis in *Solanum*. CAS cyclizes 2,3-oxidosqualene from the mevalonate pathway to form cycloartenol a common metabolite in both phytosterol and cholesterol biosynthesis. Cycloartenol is converted to campesterol by a ten-step pathway and through a nine-step pathway 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.

Published on 24 March 2022. Downloaded by Michigan State University on 10/17/2023 4:02:57 PM.

arose from duplication and divergence of the phytosterol pathway genes, C5-SD1, 7-DR1, SMO1 and SMO2.[121]

Presence-absence variation of genes involved in the conversion of dehydro-SGAs to dihydro-SGAs contributes to SGA diversity within *Solanum*. The first spirosolosane-type SGA formed, (22*S*, 25*S*)-spirosol-5-en-3β-ol, contains a $\Delta^{5,6}$ double bond.[10] In tomato, tomatidine is synthesized from a multistep process starting with the oxidation and isomerization of (22*S*, 25*S*)-spirosol-5-en-3β-ol to tomatid-4-en-3-one by GAME25, and the addition of four sugars (galactose, glucose, glucose, and xylose) to the C-3 position of tomatidine results in the production of tomatine, the major tomato SGA.[7,123,124] Lack of a functional GAME25 is associated with the production of unsaturated SGAs, including α-solamargine, α-solasonine, and malonylsolamargine in *S. melongena* (eggplant) and expression of tomato GAME25 in eggplant results in the production of saturated SGAs.[123] However, the mechanism underlying a lack of saturated SGA accumulation in domesticated potato is less clear. A putative GAME25 homolog is present in the genome of domesticated potato, and recombinant expression of the corresponding enzyme revealed the same activity as the tomato enzyme: 3β-hydroxyl group oxidation and isomerization of the double bond from the C-5,6 position. The potato GAME25 enzyme is active with unsaturated spirolosane- and solanidine-type SGAs although the corresponding saturated SGAs do not accumulate in domesticated potato.[123] Overexpression of tomato GAME25 in potato hairy root cultures leads to accumulation of demissidine, a saturated solanidine SGA found in wild potato. This suggests that the downstream enzymatic activities involved in the production of saturated SGAs exist in domesticated potato.[125] However, the mechanism leading to the lack of saturated SGAs in domesticated potato remains unclear, and the *in vivo* function of the domesticated potato GAME25 and expression levels of the corresponding gene remain to be determined.[123,125]

While the initial steps of spirolosane-type SGA formation are conserved between tomato and potato, SGA biosynthesis diverges in potato to produce solanidine-type SGAs.[10] Potato contains two major solanidane-type SGAs, α-solanine and α-chaconine, which differ only in the identity of the C-3 sugar additions; solanine contains galactose with rhamnose and glucose additions while chaconine contains glucose with two rhamnose additions.[10] The 2-oxoglutarate dependent dioxygenase, DPS (Dioxygenase for Potato Solanidane synthesis), catalyzes solanidine ring formation *via* C-16 hydroxylation.[10] While both eggplant and tomato contain DPS homologs and each recombinant enzyme is capable of C-16 hydroxylation of spirolosane-type SGAs, the expression of the corresponding genes is low or undetectable in eggplant and tomato, which likely explains the lack of solanidine-type SGAs in these species.[10] The DPS genes are located on chromosome 1 within a syntenic block that is conserved in *Solanum* and contains additional SM-related genes, suggesting that the DPS genes evolved prior to speciation.[10] While some wild potato species, such as *Solanum chacoense*, produce leptines, solanidine-type SGAs that are effective at defending against CPB, domesticated potato does not produce these SGAs. Leptine formation requires the hydroxylation of solanidine-type SGAs by GAME32 and the subsequent acetylation by an unknown enzyme. Tomato and domesticated potato lack a functional GAME32 homolog and the corresponding leptine SGAs.[117]

Domestication and selection for non-bitter fruit to aid in seed dispersal influence SGA content in tomato during fruit ripening. The fruit ripening associated biosynthesis of esculeoside A from α-tomatine alleviates the bitter taste associated with SGAs.[117] The hydroxylation of α-tomatine at the C-23 position is the first committed step of fruit ripening associated SGA accumulation (i.e. esculeoside A), and is catalyzed by the 2-ODD enzyme, GAME31.[117,126] Esculeoside A formation requires an additional hydroxylation, followed by acetylation, and the glycosylation of acetoxy-hydroxytomatine by GAME5.[117,127,128] The export of α-tomatine and α-tomatine derivatives out of the vacuole by a nitrate transporter 1/peptide transporter family (NPF) transporter, GORKY (meaning bitter in Russian), is essential for esculeoside A formation.[129] The sequestration of toxic SGAs to the vacuole likely prevents self-toxicity, and this is evidenced by the observation that tomato plants overexpressing GORKY (facilitating SGA export to the cytosol) displayed severe morphological phenotypes.[129] In contrast, fruit from the same overexpression lines did not display signs of self-toxicity suggesting that the conversion of toxic/bitter SGAs to esculeosides prevents self-toxicity.[129]

The synteny of the metabolic gene clusters involved in SGA production among *Solanum* species highlights the common origin of the trait that diverged between species through loss or gain of function of individual genes to create SGA diversity. Several of the genes involved in spirolosane-type SGA formation are found clustered on potato, eggplant, and tomato chromosomes 7 and 12.[8,122] Tomato possesses two extra genes in these clusters as potato and eggplant lack homologs of GAME17 and 18, two UDP-glucosyltransferases responsible for the consecutive additions of glucose to tomatidine galactoside during α-tomatine biosynthesis in tomato.[8] Current genomic resources show that pepper (*Capsicum annuum*) does not possess the chromosome 12 cluster or putative orthologs of GAME4 and GAME12 found within the cluster, and this absence likely results in the lack of SGAs in *C. annuum*.[122] The 2-ODD genes involved in solanidine, leptine, and esculeoside SGA biosynthesis are also clustered with additional 2-ODDs of unknown function.[117] Changes in gene expression (i.e. low expression of DPS tomato homolog) or the presence-absence of single genes (i.e. GAME32 presence in *S. chacoense*) contribute to SGA diversity in *Solanum*.

7 Addictive and deadly: convergent and divergent evolution shapes nicotine and tropane alkaloid metabolism

Several Solanaceae genera, including *Datura*, *Atropa*, *Hyoscyamus*, *Mandragora*, and *Scopolia* derive medicinal and toxic qualities from the biosynthesis of tropane alkaloids. Tropane alkaloids are characterized by an eight-membered, bicyclic,

 This journal is © The Royal Society of Chemistry 2022

Published on 24 March 2022. Downloaded by Michigan State University on 10/17/2023 4:02:57 PM.

nitrogen-containing core and their synthesis is reported in 10 plant families, separated by ~120 Mya of evolution.[130] For example, the well-known narcotic cocaine is synthesized by *Erythroxylum coca* (Erythroxylaceae) while cochlearine is synthesized in *Cochlearia officinalis* (Brassicaceae). The Solanaceae family has emerged as a model system for studying tropane alkaloid biosynthesis, but comparative studies reveal instances of independent evolution of tropanes in distinct plant lineages.[131,132]

Scopolamine and hyoscyamine are tropane aromatic esters specific to the Solanaceae, and these compounds derive their medicinal properties from anticholinergic effects, blocking activity of the neurotransmitter acetylcholine. Scopolamine is used to treat a variety of illnesses including motion sickness, drooling, and for palliative care in Parkinson's disease.[133–135] Tropane aromatic ester production requires the biosynthesis of the tropane core as well as condensation of a phenyllactic acid moiety through an ester linkage.[136] Although the biosynthesis of the tropane core intermediate and polyhydroxylated derivates, known as calystegines, occurs in many genera of the Solanaceae, including *Solanum*, the biosynthesis of tropane aromatic esters is restricted to the genera described above, suggesting that not all species in the family possess the genes required for their synthesis.[137] Due to their medicinal importance, considerable effort has focused on understanding the biosynthesis of hyoscyamine and scopolamine.

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

The first ring of the tropane core requires the conversion of ornithine, a non-proteinogenic amino acid, into putrescine by ornithine decarboxylase (ODC). Putrescine is then *N*-methylated by putrescine methyltransferase (PMT) and oxidized by methylputrescine oxidase (MPO). The *N*-methyl-Δ^1-pyrrolinium cation forms through the spontaneous cyclization of *N*-methylaminobutanal, the product of MPO catalysis (Fig. 7). PMT requires *S*-adenosyl-L-methionine (SAM) to *N*-methylate putrescine and shares high sequence similarity with spermidine synthase (*SPDS*), an enzyme involved in transferring the aminopropyl moiety from decarboxylated SAM (dcSAM) onto putrescine to form spermidine, a ubiquitous polyamine.[139,140] It was hypothesized that *PMT* evolved from a gene duplication of *SPDS* and subsequent neofunctionalization, and although SPDS cannot catalyze putrescine *N*-methylation, mutation of a single SPDS amino acid, D103I, is sufficient to generate PMT activity.[139] The pyrrole moiety of nicotine, a natural product produced in the *Nicotiana* genus of the Solanaceae, also requires *N*-methyl-Δ^1-pyrrolinium cation biosynthesis. The biosynthetic steps leading to *N*-methyl-Δ^1-pyrrolinium cation formation are conserved in *Nicotiana*, *Solanum*, and *Petunia* allowing the *N*-methyl-Δ^1-pyrrolinium cation to act as a core for nicotine and tropane alkaloid biosynthesis found in Solanaceae and Convolvulaceae.[141,142] In contrast, the genes involved in the formation of the pyridine ring in nicotine biosynthesis are *Nicotiana*-specific indicating that divergent evolution led to the formation of nicotine, likely through the duplication of the genes in the nicotinamide adenine dinucleotide (NAD) cofactor biosynthetic pathway.[142]

Formation of the tropane core in Solanaceae species requires a second cyclization event that yields tropinone, which possesses a ketone functional group at the carbon-3 position of the core (Fig. 7). The first step in tropinone formation is catalyzed by a type III polyketide synthase, PYKS, which uses the *N*-methyl-Δ^1-pyrrolinium cation and malonyl-Coenzyme A to form 4-(1-methyl-2-pyrrolidinyl)-3-oxobutanoic acid.[143] Although PYKS can form 3-oxoglutaric acid without the *N*-methyl-Δ^1-pyrrolinium cation and these two products can react non-enzymatically, the exact mechanism of 4-(1-methyl-2-pyrrolidinyl)-3-oxobutanoic acid formation remains unclear.[144,145] Tropinone synthase (CYP82M3) converts 4-(1-methyl-2-pyrrolidinyl)-3-oxobutanoic acid to tropinone[143] Although putative orthologs of PYKS and CYP82M3 are present in the genomes of several calystegine producing Solanaceous species including tomato, potato, and pepper, these genes are absent in *Nicotiana* spp.; this is consistent with the lack of detectable tropanes in these species.[143] In the Solanaceae, tropinone reductases I and II are members of the short-chain dehydrogenase/reductase superfamily (SDR) that catalyze the reduction of the ketone of tropinone to an alcohol to form tropine (3α-hydroxytropine) and pseudotropine (3β-hydroxytropine), respectively.[146] TRI and TRII constitute a branch point in the tropane alkaloid biosynthetic pathway due to their stereospecificity: TRI leads to the production of tropane aromatic esters, including hyoscyamine and scopolamine and TRII directs flux towards calystegine production.

Biosynthesis of the principal aromatic tropane esters in the Solanaceae, littorine, hyoscyamine, and scopolamine, requires the diversion of phenylalanine into the tropane pathway through a two-step process that yields phenyllactic

Published on 24 March 2022. Downloaded by Michigan State University on 10/17/2023 4:02:57 PM.

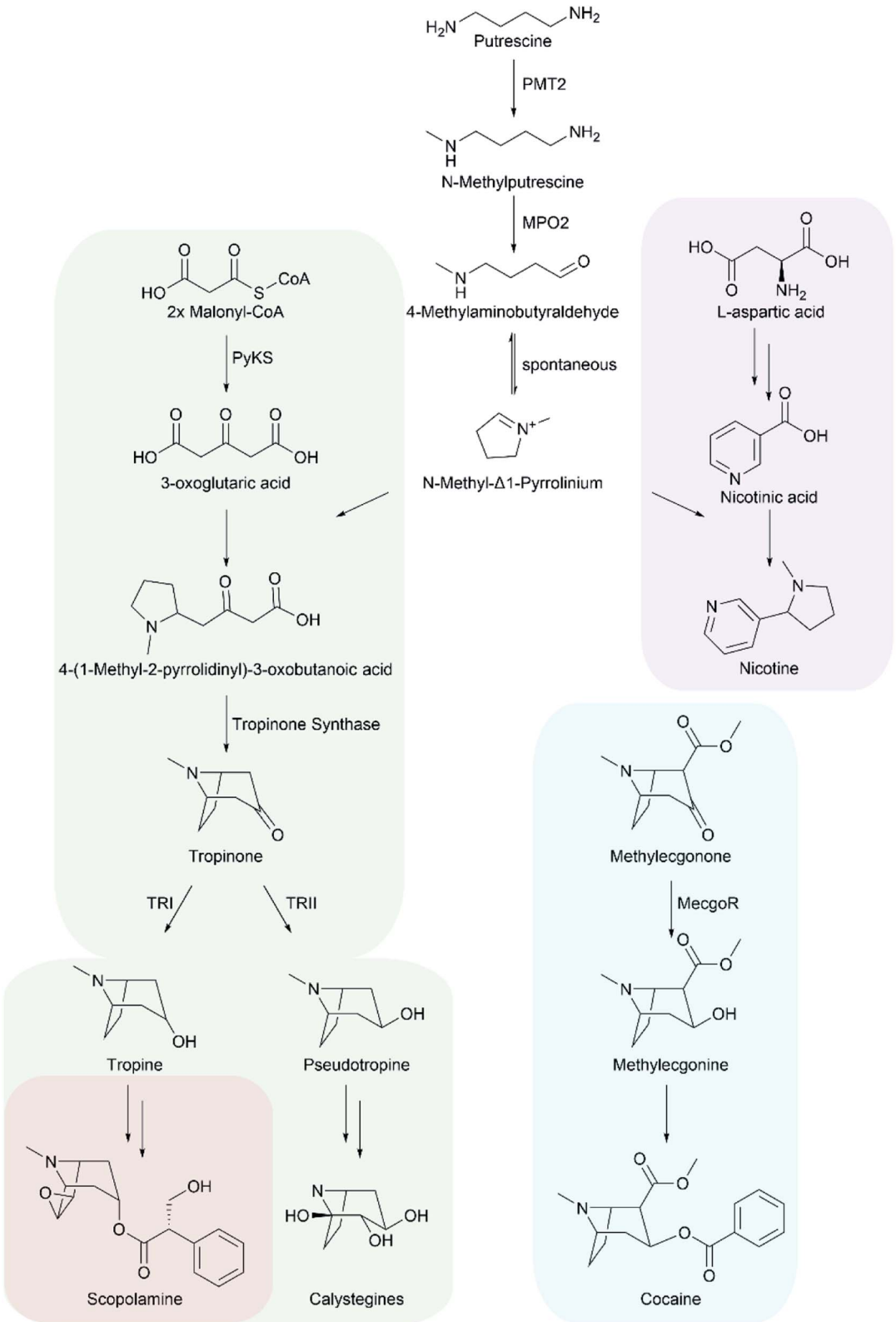


Fig. 7 Evolutionary trajectories of tropane and nicotine formation in distinct plant lineages. Comparison of tropane and nicotine alkaloid biosynthesis reveals examples of both convergent (cocaine biosynthesis in *E. coca*) and divergent (nicotine biosynthesis) evolution.[132,142] Scopolamine (orange) 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.

 This journal is © The Royal Society of Chemistry 2022

Published on 24 March 2022. Downloaded by Michigan State University on 10/17/2023 4:02:57 PM.

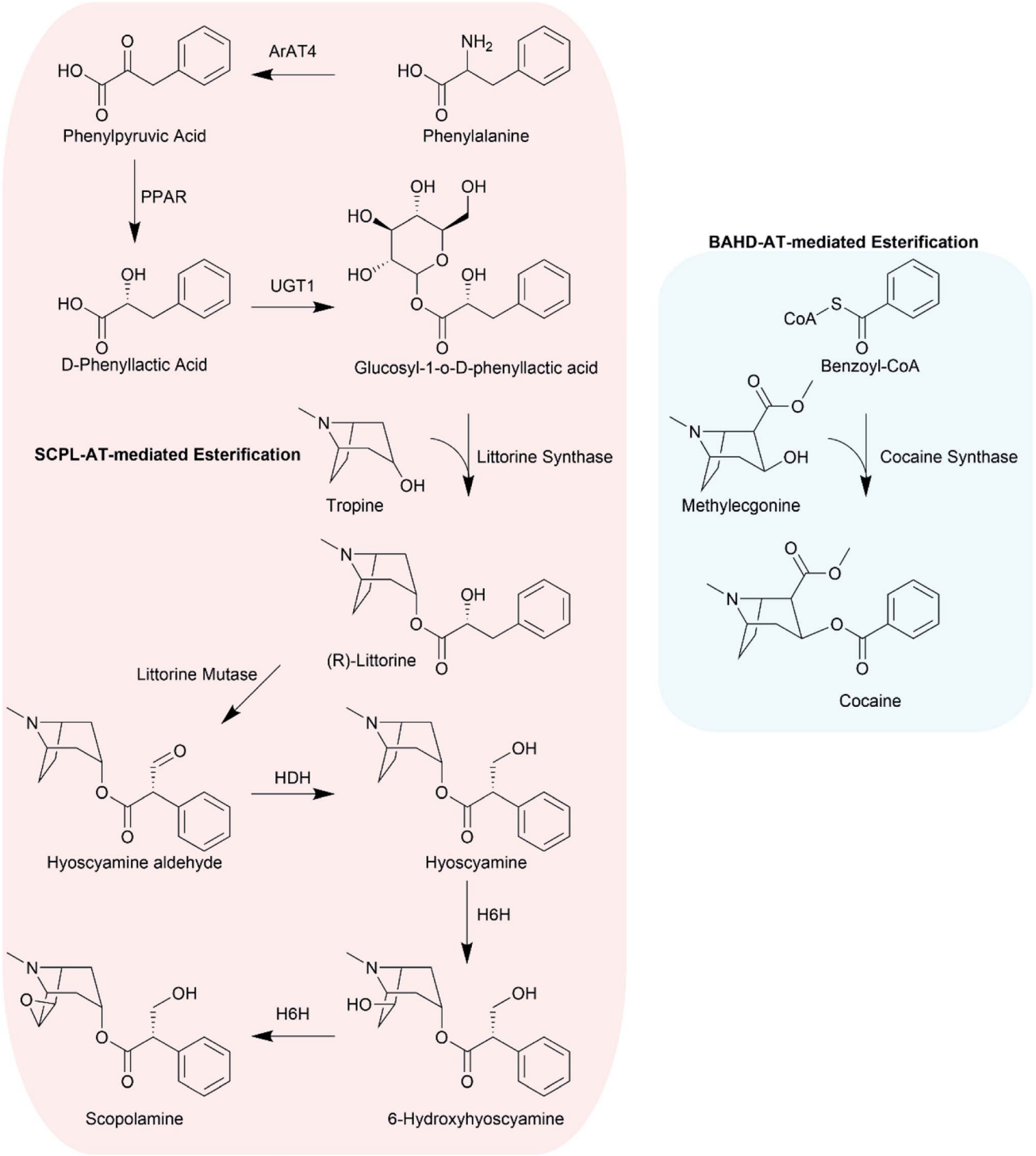


Fig. 8 Independent evolution of tropane aromatic ester formation in Solanaceae and Erythroxylaceae. Scopolamine biosynthesis requires the biosynthesis of D-phenyllactic acid *via* a two-step process mediated by ArAT4 and PPAR.[147,148] D-Phenyllactic acid is glycosylated by UGT1 to form a glucose ester 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.

Published on 24 March 2022. Downloaded by Michigan State University on 10/17/2023 4:02:57 PM.

acid[147,148](Fig. 8). Identification of the aromatic aminotransferase (AbArAT4) responsible for conversion of phenylalanine into phenylpyruvate revealed the power of transcriptomics in Solanaceae tropane alkaloid enzyme discovery.[147] Analogous to bacterial aromatic amino acid biosynthesis, a cytosolic aromatic aminotransferase from petunia (Ph-PPY-AT) catalyzes the formation of phenylalanine from phenylpyruvate using tyrosine as an amino donor and yielding 4-hydroxyphenylpyruvate.[149] AbArAT4 is related to Ph-PPY-AT and utilizes the same four substrates, but the Atropa enzyme diverts phenylalanine into the tropane pathway by virtue of a ~250-fold more active reverse reaction that yields phenylpyruvate and tyrosine. *AbArAT4* is co-expressed in the roots with other tropane-related genes, and while silencing of this gene disrupts tropane alkaloid biosynthesis, it does not alter aromatic amino acid pools, further supporting its neofunctionalized and specific role in specialized metabolism.[147] Littorine biosynthesis requires the glycosylation of phenyllactate by a UDP-glucose dependent glycosyltransferase followed by the acylation of tropine. The serine carboxypeptidase-like (SCPL) acyltransferase (littorine synthase) acylates tropine using glycosylated phenyllactate as the acyl donor.[136]

Synthetic biology recently was utilized both to engineer scopolamine production in yeast and facilitate the discovery of the final missing enzyme in the pathway, which had eluded discovery using *in planta* experiments. The conversion of littorine to scopolamine requires four steps catalyzed by three enzymes (Fig. 8). Littorine mutase, a cytochrome P450, catalyzes the rearrangement of littorine to hyoscyamine aldehyde,[138] which is converted to hyoscyamine by hyoscyamine aldehyde dehydrogenase. Finally, hyoscyamine-6-hydroxylase catalyzes the two-step hydroxylation and epoxidation of hyoscyamine to scopolamine.[150] The production of scopolamine in yeast was achieved through the introduction of tropane alkaloid pathway genes from several species, including *Datura stramonium*, *Datura metel*, and *Atropa belladonna*.[151] Optimization of scopolamine production in yeast required the elimination of several native genes to reduce the flow of tropane alkaloid intermediates into side products and the introduction of a transporter from *Nicotiana tabacum* to facilitate transport of tropine into the vacuole for esterification with phenyllactic acid.[151] Notably, the introduction of the pathway into yeast revealed the dehydrogenase responsible for the reduction of hyoscyamine aldehyde into hyoscyamine, which had not previously been identified *in planta*.[151] For example, silencing of this gene in *A. belladonna* did not result in a decrease in downstream tropane alkaloids, likely due to promiscuous enzymatic activity of other dehydrogenases.[152] Hence, reconstruction of the pathway in a genetic host where background activities were removed facilitated the identification of the final missing step in the scopolamine pathway.

7.1 Independent evolution of tropanes in distinct plant lineages

Evidence for independent evolution of tropanes in distinct plant lineages is manifest at different steps throughout the pathway (Fig. 7 and 8). While separate TRI and TRII enzymes reduce tropinone to tropine or pseudotropine in the Solanaceae, a single SDR enzyme catalyzes both reactions in *C. officinalis*, ultimately leading to tropine-derived cochlearine and pseudotropine-derived calystegines.[131] In addition, while Solanaceae and Brassicaceae species utilize enzymes in the SDR family for the reduction of tropinone, the analogous reaction in *E. coca* cocaine biosynthesis, the reduction of methylecgonone to methylecgonine, is catalyzed by methylecgonone reductase (MecgoR) a member of the aldo-keto reductase family.[132] Similarly, aromatic tropane ester biosynthesis is catalyzed by different classes of acyltransferases in the Solanaceae and Erythroxylaceae. Littorine formation is synthesized by an SCPL acyltransferase while cocaine synthase, which catalyzes the condensation of methylecgonine and benzoyl-CoA, is a member of the BAHD acyltransferase family.[153] As additional tropane pathways in distinct plant lineages are elucidated it is likely that further examples of independent evolution will be discovered.

8 Challenges and unexplored frontiers in Solanaceae metabolism

There has been a rapid increase in understanding the biosynthesis and evolution of plant SM pathways during the last decade. Advances in genomics enabled gene–metabolite correlations in model and non-model species. These data – combined with development of methods to test gene function in diverse species, and transient expression in *Nicotiana benthamiana*, as well as engineering production in microbial systems – led to the elucidation of multiple plant SM pathways and identified regulators of known SM pathways.[2–4,151,154] The widespread adoption of these approaches, coupled with phylogeny-guided comparative genomics and metabolomics, enabled exploration of the evolutionary trajectories of the exemplary Solanaceae SM pathways described here.

However, despite advances in understanding Solanaceae SM biosynthesis and evolution, knowledge gaps persist related to specific aspects of these well-studied pathways and opportunities exist to develop a more comprehensive understanding of these pathways and networks. As evidenced through studies of acylsugar evolution, much can be learned through adopting a broader sampling strategy to include more phylogenetically diverse species that are typically less well studied.[37,72,78] Similar, phylogenetic-guided metabolite screening approaches could be adopted to assess chemical diversity in other SM classes as the foundation for exploring metabolite evolution using comparative genomics. For example, given the tremendous chemical variation observed in trichome-derived acylsugars across the Solanaceae, and that novel acylsugars were recently identified in root and root-exudates of tomato,[48] it will be intriguing to determine whether comparable root acylsugar diversity exists across the family and if so, to assess how this diversity evolved.

There are also several examples where the biosynthesis of exemplary SM pathways in the Solanaceae are not fully resolved. For example, the enzymes that catalyze the early steps in acylinositol biosynthesis in *Solanum* spp. are yet to be reported.

 This journal is © The Royal Society of Chemistry 2022

Published on 24 March 2022. Downloaded by Michigan State University on 10/17/2023 4:02:57 PM.

Similarly, the majority of the enzymes involved in capsaicinoid biosynthesis and the final steps in nicotine biosynthesis await biochemical and functional characterization.[142,155] In addition, although the biosynthesis of scopolamine is elucidated and the pathway reconstructed in yeast, the steps leading to the biosynthesis of other classes of Solanaceae tropanes, including calystegines and schizanthines, are unknown.[130,156]

Comparative analyses of the evolution of SM-related gene clusters across the Solanaceae also remains under-explored. For example, as outlined in this review, terpene and SGA-related gene clusters exist in *Solanum* but variation across these clusters is mainly documented in a few model species, including tomato, potato, eggplant, and closely related wild species.[8,82,122] Indeed, even for the comparatively well-studied terpenoid-related gene clusters of tomato, many of the enzymes that reside within these clusters, which may catalyze modifications of terpene scaffolds, remain uncharacterized. Furthermore, the extent of conservation of terpene and other SM gene clusters across the Solanaceae is unknown. As multiple chromosome scale genome assemblies of phylogenetically diverse Solanaceae species are available and others will likely be generated soon, charting the evolutionary trajectories of SM gene clusters and the metabolite variation they encode is now possible.

Finally, it is also worth noting that the most extensively characterized Solanaceae SM pathways are those where the identities of the major metabolites were known for decades and their abundance is high in specific cell types or tissues, facilitating purification and structural elucidation. It is more challenging to identify unknown metabolites and purify metabolites that are of low abundance and technical challenges persist that impede a more comprehensive understanding of metabolism and bridging of the gap between genotype and phenotype.

8.1 Challenges in the identification and annotation of SM enzymes

Advances in DNA sequencing are making development of chromosome-scale genome assemblies more routine and recently several Solanaceae genomes were released, and the quality of existing assemblies improved.[17,19,157] These studies allow the gene complement of an organism to be determined. However, functional annotation of plant genomes remains incomplete, even for model species. The lack of accurate annotation is particularly problematic for large gene families encoding SM-related enzymes that catalyze common decorations of scaffold molecules, including cytochromes P450, 2-oxoglutarate dependent dioxygenases, glycosyltransferases, and acyltransferases. SM-related enzymes are often catalytically promiscuous and encoded by genes that evolved rapidly through duplication and associated subfunctionalization, neofunctionalization, and gene loss.[158] Thus, annotation of SM enzymes based solely on sequence similarity, predicted orthology, or synteny is often misleading. This concept is clearly illustrated by examples identified through studying the evolution of acylsugar and terpene biosynthesis in *Solanum* glandular trichomes. These studies reveal how activity can be altered by a few amino acid differences in closely related enzymes from sister species, or diverse accessions within a species.[64,90,95] Hence, empirical determination of enzyme function remains imperative. Although characterization of enzyme activities is often technically challenging, time consuming, and limited by substrate availability, medium and high-throughput methods based on microtiter plates and microfluidics are utilized for screening natural and computationally designed enzymes and such methods could potentially be adapted for screening the activity of plant SM-related enzymes.[159]

As documented throughout this review, co-expression is a powerful approach for predicting membership of genes in metabolic pathways, particularly when there is *a priori* knowledge about enzymes from the target pathway. Elucidation of the pathway leading to scopolamine biosynthesis, described above, is an excellent example of the use of co-expression analyses to identify candidate genes co-expressed in roots. However, when results of co-expression analysis are ambiguous or multiple candidate genes are identified, as is often the case when investigating large SM-related gene families, additional filtering and refinement of gene candidates may be required prior to time-consuming functional studies. In such cases, comparative genomic analysis such as synteny or gene-cluster analysis – together with phylogenetic analysis to determine whether gene candidates exhibit lineage-specific distribution or arose through a recent duplication event – provide opportunities for refining candidate gene lists.[160] Outside of tomato, there is a lack of publicly available transcriptome data, including data from diverse tissues, environmental perturbations, and treatments. This limits novel metabolite pathway discovery in diverse Solanaceae species and reduces the resolution of studies investigating the phylogenetic distribution and evolution of SM pathways. Furthermore, plant SM pathways are often restricted to specific cell types, and therefore the general focus on whole tissue sampling for transcriptome analysis can be limiting.[68,161,162] The recent development of single-cell and single-nucleus transcriptome analyses holds great promise for increasing the resolution of transcriptome data and refining candidate gene lists to facilitate the identification, characterization, and cellular localization of Solanaceae SM pathways.[163,164]

Machine learning is another promising approach to distinguish GM and SM-related enzymes without prior knowledge of pathway membership or gene–metabolite correlation information. Multiple features including gene expression, transcriptional network analysis, rate of evolution, and duplication mechanism allowed creation of statistical models that can distinguish GM from SM genes in *Arabidopsis*. In agreement with the established characteristics of SM genes, machine learning models revealed that relative to GM genes, SM genes tend to be less conserved, tandemly duplicated, more narrowly expressed, and expressed at lower levels.[165] The prediction models also facilitated the classification of 1220 enzyme encoding genes of unknown function as putatively SM-related. Similar machine learning strategies were deployed in tomato to predict gene association with SM or GM pathways and to determine if gene expression data can predict metabolic pathway membership.[166,167] These approaches show potential to

Published on 24 March 2022. Downloaded by Michigan State University on 10/17/2023 4:02:57 PM.

build high-quality models but are limited by the quality of the input data, including mis-annotations and the low number of functionally validated reference genes in tomato. These current limitations suggest that application of machine learning for *de novo* prediction of novel SM pathways in tomato is not yet possible at high accuracy. Furthermore, additional functional annotation, including the development of more comprehensive genome and transcriptome data, will be needed to apply machine learning approaches to predict SM pathway membership in additional members of the Solanaceae. Indeed, models predicting whether a tomato gene is associated with specialized *versus* general metabolism were improved when a transfer learning strategy was employed that utilized data from *Arabidopsis* models to filter tomato annotations that disagreed with *Arabidopsis*[166] This represents a promising approach to using comparative genomics data in specialized metabolic enzyme identification.

8.2 Challenges in the identification and annotation of plant metabolites

Estimates suggest that $\sim 10^6$ metabolites are synthesized across species of the plant kingdom, collectively.[1] While we have deep knowledge of well-studied classes of plant metabolites, opportunities and challenges for improving metabolome annotation remain. Several factors make separation and annotation of metabolites challenging: for example, their diverse chemical composition, chemical properties (polarity and hydrophilicity/hydrophobicity), and the orders of magnitude concentration range in which they occur in biological samples.[168,169] Improvements in analytical techniques, particularly liquid-chromatography coupled with high-resolution mass-spectrometry (LC-HRMS) based metabolite profiling, allows the detection of $>10^3$ metabolites within a single plant extract at high mass accuracy. However, a single extraction solvent and chromatographic separation method are generally selected for individual experiments, leading to unavoidable bias in the types of metabolites that are extracted and resolved and therefore an under-representation of the metabolome.[168] Furthermore, most metabolites in a plant extract are uncharacterized and many are of low abundance. In such cases, annotation can be challenging. This is particularly true for specialized metabolites that are formed from diverse metabolic precursors, possess multiple chemical modifications, and frequently exist as positional or structural isomers that may be difficult to resolve. For example, even though tomato fruit ripening is one of the most extensively studied plant biological processes, a large component of this metabolome remains unannotated. In a recent study, untargeted metabolomics of tomato fruit at two different developmental stages identified >1000 semi-lipophilic metabolites but only $\sim$170 metabolites were annotated with some degree of confidence, suggesting that the bulk of the tomato fruit metabolome remains unresolved.[127] Metabolite databases containing spectra derived from tandem mass-spectrometry of known metabolites are expanding and are useful for identifying unknown metabolites.[170–172] However, given the vast diversity of plant metabolites and their frequent lineage-specific distribution, populating and curating such databases requires substantial research funding, effort, and community engagement.

As with spatially resolved or single cell transcriptomics, the ability to obtain spatially resolved metabolome data through mass spectrometry imaging of plant tissues represents an exciting development that will enhance understanding of metabolism. Specifically, this technology will further refine the ability to detect gene–metabolite correlations and allow the detection of metabolites that may be restricted to individual cell types and therefore fall below the limit of detection in an extract prepared from a complex tissue sample.[173] Mass spectrometry imaging has been utilized for investigating the spatial distribution of metabolites in tomato fruit, including investigating the influence of genetic perturbation on SGA accumulation.[174] Similarly, the spatial separation of SGAs and acylsugars were demonstrated in tomato roots.[48] As improved MSI technologies develop and increase in availability, they will undoubtedly be more widely adopted for exploring diverse aspects of Solanaceae metabolism.

Integration of genetic variation with metabolomics is a powerful approach to expand understanding of SM metabolic networks and bridge the gap between genotype and phenotype. As described above, both GWAS and metabolite QTL (mQTL) approaches were used to identify genomic regions and genes that influence specialized metabolism in diverse tissues of tomato. In particular, the *S. lycopersicum* x *S. pennellii* introgression line and the related backcross introgression line (BIL) populations were foundational to improving understanding of the loci that influence metabolism within the tomato clade.[33,62,117,127,175] Approaches that harness natural variation are limited to species where it is possible to develop inter-specific genetic populations or sufficient genetic variation is present within a species, to facilitate GWAS. Although not currently as extensively characterized as the genetic resources for tomato, germplasm panels and genetic populations, including introgression lines, are being developed and characterized for the three additional major food crops of the Solanaceae; potato, pepper, and eggplant.[105,176,177] In some cases, these genetic resources are being utilized to investigate metabolic diversity *via* targeted and untargeted metabolomics and refinement of these efforts should facilitate linking genotype to phenotype.[178,179]

An alternative, less frequently utilized, approach to harness genetic variation to interrogate metabolism is to combine untargeted metabolite profiling with targeted disruption or over-expression of known enzymes or transcription factors.[180,181] This approach, while more targeted than a strategy incorporating genome-wide genetic variation, can be utilized in any species where genetic manipulation is feasible and has significant potential to increase understanding of plant SM networks. For example, disruption of an SM enzyme will result in reduction of metabolites downstream of the enzyme, while the abundance of metabolites upstream of the target enzyme can increase. This approach also allows detection of alternate fates for pathway metabolites that accumulate due to gene disruption, revealing the existence of biosynthetically linked

 This journal is © The Royal Society of Chemistry 2022

Published on 24 March 2022. Downloaded by Michigan State University on 10/17/2023 4:02:57 PM.

metabolites. Referred to as "silent metabolism" this component of the metabolome is likely substantial and certainly underexplored, including for engineering of novel products.[182] Furthermore, as SM enzymes possess increased tendency for catalytic promiscuity, untargeted metabolite profiling of lines disrupted in an enzyme of interest may reveal the existence of previously uncharacterized catalytic activities.

While purification and structural elucidation of metabolites by NMR is a cornerstone of SM pathway discovery, it is time-consuming and typically represents a major bottleneck. This is especially problematic for metabolites that are of low abundance or co-purify with other compounds. Recent structural elucidation of acyl-hexoses from *S. nigrum* was achieved using a combination of LC-MS, GC-MS, and 2D-NMR approaches from crude and partially purified extracts without purification to homogeneity.[72] Similar approaches should be adaptable to resolve the structures of other metabolites present in semi-purified plant extracts. The recent adoption of microcrystal electron diffraction (MicroED) for structural elucidation, including absolute stereochemistry, of mixtures of small organic molecules also shows great promise for structural elucidation of plant specialized metabolites.[183,184] MicroED can be used to resolve the structures of nanocrystals of ~100 nm (~10^{-15} g) and thus is potentially more suitable for low abundance metabolites than NMR, which typically requires hundreds of micrograms to milligram quantities of purified compound. Application of this technology to specialized metabolite discovery was recently demonstrated through a combined genome-mining, synthetic biology, and MicroED analysis that elucidated the biosynthesis and structures of several 2-pyrridone metabolites from fungi.[185] Similarly, synthetic biology can be utilized to engineer production of plant SMs in heterologous systems for subsequent purification and structural elucidation. This strategy was effectively demonstrated by the synthesis of gram scale quantities of the triterpene β-amyrin by vacuum infiltration of *N. benthamiana* co-expressing a feedback insensitive variant of HMG-CoA reductase and oat β-amyrin synthase.[186] Subsequent experiments combining co-expression of these enzymes with triterpene decorating cytochrome P450s from multiple species facilitated the production of novel non-natural triterpenes at sufficient scale to allow purification and structural determination by NMR. *N. benthamiana* is widely used for transient expression of candidate genes and as demonstrated above, represents a readily scalable platform to produce metabolites for purification and subsequent structural elucidation.

9 Conclusions

Advances in genomics and metabolomics continue to enable greater understanding of SM pathway biosynthesis and evolution. This review focused on the catalytic steps of five well-studied SM classes that show varying degrees of lineage-specific distribution across the Solanaceae. This genetic variation, coupled with high abundance, and often restricted distribution in specific tissue or cell types, facilitated both purification and structural elucidation of these diverse metabolites as well as the identification of the enzymes responsible for their biosynthesis. For example, acylsugar and terpene biosynthesis in glandular trichomes, nicotine and tropane alkaloid biosynthesis in roots, and capsaicinoid biosynthesis in pepper fruit placenta. These studies reveal examples of both intra- and inter-specific variation as well as convergent evolution that has shaped the metabolic landscape across the Solanaceae. However, only a small fraction of the metabolome and the genes responsible for its formation are resolved. Thus, many opportunities exist to expand understanding of known pathways as well as identify novel pathways that will enable a network level understanding of metabolism across the Solanaceae and identify target molecules for agricultural and medicinal applications.

10 Conflicts of interest

There are no conflicts to declare.

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 Cell Physiol.*, 2012, **53**, e1.
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, *Nat. Commun.*, 2014, **5**, 3283.
5 T. Vogt, *Mol. Plant*, 2010, **3**, 2–20.
6 C. A. Schenck and R. L. Last, *FEBS J.*, 2020, **287**, 1359–1368.
7 P. D. Sonawane, A. Jozwiak, S. Panda and A. Aharoni, *Curr. Opin. Plant Biol.*, 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.

Published on 24 March 2022. Downloaded by Michigan State University on 10/17/2023 4:02:57 PM.

9 B. J. Leong and R. L. Last, *Curr. Opin. Struct. Biol.*, 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, *Nat. Commun.*, 2021, **12**, 1300.
11 T. Särkinen, L. Bohs, R. G. Olmstead and S. Knapp, *BMC Evol. Biol.*, 2013, **13**, 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, *Am. J. Bot.*, 2022, DOI: 10.1002/ajb2.1827.
13 S. Knapp, L. Bohs, M. Nee and D. M. Spooner, *Comp. Funct. 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, E. van der Knaap, S. Huang, H. J. Klee, J. J. Giovannoni and Z. Fei, *Nat. Genet.*, 2019, **51**, 1044–1051.
17 M. Alonge, X. Wang, M. Benoit, S. Soyk, L. Pereira, L. Zhang, H. Suresh, S. Ramakrishnan, F. Maumus, D. Ciren, Y. Levy, T. H. Harel, G. Shalev-Schlosser, Z. Amsellem, H. Razifard, A. L. Caicedo, D. M. Tieman, H. Klee, M. Kirsche, S. Aganezov, T. R. Ranallo-Benavidez, Z. H. Lemmon, J. Kim, G. Robitaille, M. Kramer, S. Goodwin, W. R. McCombie, S. Hutton, J. Van Eck, J. Gillis, Y. Eshed, F. J. Sedlazeck, E. van der Knaap, M. C. Schatz and Z. B. Lippman, *Cell*, 2020, **182**, 145–161.
18 L. A. Mueller, T. H. Solow, N. Taylor, B. Skwarecki, R. Buels, J. Binns, C. Lin, M. H. Wright, R. Ahrens, Y. Wang, E. V. Herbst, E. R. Keyder, N. Menda, D. Zamir and S. D. Tanksley, *Plant Physiol.*, 2005, **138**, 1310–1317.
19 L. Barchi, M. T. Rabanus-Wallace, J. Prohens, L. Toppino, S. Padmarasu, E. Portis, G. L. Rotino, N. Stein, S. Lanteri and G. Giuliano, *Plant Journal*, 2021, **107**, 579–596.
20 B. Song, Y. Song, Y. Fu, E. B. Kizito, S. N. Kamenya, P. N. Kabod, H. Liu, S. Muthemba, R. Kariba, J. Njuguna, S. Maina, F. Stomeo, A. Djikeng, P. S. Hendre, X. Chen, W. Chen, X. Li, W. Sun, S. Wang, S. Cheng, A. Muchugi, R. Jamnadass, H.-Y. Shapiro, A. Van Deynze, H. Yang, J. Wang, X. Xu, D. A. Odeny and X. Liu, *Gigascience*, 2019, **8**, giz115.
21 Y. Eshed and D. Zamir, *Genetics*, 1995, **141**, 1147–1162.
22 I. Ofner, J. Lashbrooke, T. Pleban, A. Aharoni and D. Zamir, *Plant Journal*, 2016, **87**, 151–160.
23 T. W. Toal, M. Ron, D. Gibson, K. Kajala, B. Splitt, L. S. Johnson, N. D. Miller, R. Slovak, A. Gaudinier, R. Patel, M. de Lucas, N. J. Provart, E. P. Spalding, W. Busch, D. J. Kliebenstein and S. M. Brady, *G3: Genes, Genomes, Genet.*, 2018, **8**, 3841–3855.
24 A. L. Schilmiller, A. L. Charbonneau and R. L. Last, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 16377–16382.
25 E. Fridman, F. Carrari, Y.-S. Liu, A. R. Fernie and D. Zamir, *Science*, 2004, **305**, 1786–1789.
26 C. Brooks, V. Nekrasov, Z. B. Lippman and J. Van Eck, *Plant Physiol.*, 2014, **166**, 1292–1297.
27 Y. Liu, M. Schiff and S. P. Dinesh-Kumar, *Plant Journal*, 2002, **31**, 777–786.
28 E. G. W. M. Schijlen, C. H. R. de Vos, S. Martens, H. H. Jonker, F. M. Rosin, J. W. Molthoff, Y. M. Tikunov, G. C. Angenent, A. J. van Tunen and A. G. Bovy, *Plant Physiol.*, 2007, **144**, 1520–1530.
29 J. Van Eck, *Curr. Opin. Biotechnol.*, 2018, **49**, 35–41.
30 D. Tieman, G. Zhu, M. F. R. Resende Jr, T. Lin, M. Taylor, B. Zhang, H. Ikeda, Z. Liu, J. Fisher, I. Zemach, A. Monforte, D. Zamir, A. Granell, M. Kirst, S. Huang, H. Klee, C. Nguyen, D. Bies, J. L. Rambla and K. S. O. Beltran, *Science*, 2017, **355**, 391–394.
31 G. Zhu, S. Wang, Z. Huang, S. Zhang, Q. Liao, C. Zhang, T. Lin, M. Qin, M. Peng, C. Yang, X. Cao, X. Han, X. Wang, E. van der Knaap, Z. Zhang, X. Cui, H. Klee, A. R. Fernie, J. Luo and S. Huang, *Cell*, 2018, **172**, 249–261.
32 X. Li, D. Tieman, Z. Liu, K. Chen and H. J. Klee, *Plant Journal*, 2020, **104**, 631–644.
33 K. Garbowicz, Z. Liu, S. Alseekh, D. Tieman, M. Taylor, A. Kuhalskaya, I. Ofner, D. Zamir, H. J. Klee, A. R. Fernie and Y. Brotman, *Mol. Plant*, 2018, **11**, 1147–1165.
34 M. I. Zanor, J.-L. Rambla, J. Chaib, A. Steppa, A. Medina, A. Granell, A. R. Fernie and M. Causse, *J. Exp. Bot.*, 2009, **60**, 2139–2154.
35 Y. M. Tikunov, J. Molthoff, R. C. H. de Vos, J. Beekwilder, A. van Houwelingen, J. J. J. van der Hooft, M. Nijenhuis-de Vries, C. W. Labrie, W. Verkerke, H. van de Geest, M. V. Zamora, S. Presa, J. Luis Rambla, A. Granell, R. D. Hall and A. G. Bovy, *Plant Cell*, 2013, **25**, 3067–3078.
36 E. Maldonado, F. R. Torres, M. Martínez and A. L. Pérez-Castorena, *J. Nat. Prod.*, 2006, **69**, 1511–1513.
37 G. D. Moghe, B. J. Leong, S. M. Hurney, A. D. Jones and R. L. Last, *eLife*, 2017, **6**, e28468.
38 T. Asai and Y. Fujimoto, *Phytochemistry*, 2010, **71**, 1410–1417.
39 T. Asai, N. Hara and Y. Fujimoto, *Phytochemistry*, 2010, **71**, 877–894.
40 T. Asai and Y. Fujimoto, *Phytochem. Lett.*, 2011, **4**, 38–42.
41 T. Asai, T. Sakai, K. Ohyama and Y. Fujimoto, *Chem. Pharm. Bull.*, 2011, **59**, 747–752.
42 M. Bah and R. Pereda-Miranda, *Tetrahedron*, 1996, **52**, 13063–13080.
43 R. Pereda-Miranda, R. Mata, A. L. Anaya, D. B. M. Wickramaratne, J. M. Pezzuto and A. D. Kinghorn, *J. Nat. Prod.*, 1993, **56**, 571–582.
44 M. Ono, A. Takigawa, H. Muto, K. Kabata, M. Okawa, J. Kinjo, K. Yokomizo, H. Yoshimitsu and T. Nohara, *Chem. Pharm. Bull.*, 2015, **63**, 641–648.
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.-S. Jeong, E.-M. Ahn, G.-S. Kim and N.-I. Baek, *Carbohydr. Res.*, 2013, **372**, 9–14.
46 C. Li, Z. Wang and A. D. Jones, *Anal. Bioanal. Chem.*, 2014, **406**, 171–182.

 This journal is © The Royal Society of Chemistry 2022

Published on 24 March 2022. Downloaded by Michigan State University on 10/17/2023 4:02:57 PM.

47 T. Nakashima, H. Wada, S. Morita, R. Erra-Balsells, K. Hiraoka and H. Nonami, *Anal. Chem.*, 2016, **88**, 3049–3057.
48 E. Korenblum, Y. Dong, J. Szymanski, S. Panda, A. Jozwiak, H. Massalha, S. Meir, I. Rogachev and A. Aharoni, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, **117**, 3874–3883.
49 B. Ghosh, T. C. Westbrook and A. D. Jones, *Metabolomics*, 2014, **10**, 496–507.
50 X. Liu, M. Enright, C. S. Barry and A. D. Jones, *Metabolomics*, 2017, **13**, 85.
51 C.-R. Zhang, W. Khan, J. Bakht and M. G. Nair, *Food Chem.*, 2016, **196**, 726–732.
52 C.-Y. Zhang, J.-G. Luo, R.-H. Liu, R. Lin, M.-H. Yang and L.-Y. Kong, *Fitoterapia*, 2016, **114**, 138–143.
53 D. B. Lybrand, T. M. Anthony, A. D. Jones and R. L. Last, *Metabolites*, 2020, **10**, 401.
54 E. Cicchetti, L. Duroure, E. Le Borgne and R. Laville, *J. Nat. Prod.*, 2018, **81**, 1946–1955.
55 C.-M. Cao, X. Wu, K. Kindscher, L. Xu and B. N. Timmermann, *J. Nat. Prod.*, 2015, **78**, 2488–2493.
56 C.-A. Bernal, L. Castellanos, D. M. Aragón, D. Martínez-Matamoros, C. Jiménez, Y. Baena and F. A. Ramos, *Carbohydr. Res.*, 2018, **461**, 4–10.
57 S. M. Hurney, *PhD*, Michigan State University, 2018.
58 B. M. Leckie, D. A. D'Ambrosio, T. M. Chappell, R. Halitschke, D. M. De Jong, A. Kessler, G. G. Kennedy and M. A. Mutschler, *PLoS One*, 2016, **11**, e0153345.
59 J. C. Goffreda, M. A. Mutschler, D. A. Avé, W. M. Tingey and J. C. Steffens, *J. Chem. Ecol.*, 1989, **15**, 2135–2147.
60 L. Van Thi, A. Weinhold, C. Ullah, S. Dressel, M. Schoettner, K. Gase, E. Gaquerel, S. Xu and I. T. Baldwin, *Plant Physiol.*, 2017, **174**, 370–386.
61 A. Weinhold and I. T. Baldwin, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 7855–7859.
62 A. Schilmiller, F. Shi, J. Kim, A. L. Charbonneau, D. Holmes, A. D. Jones and R. L. Last, *Plant Journal*, 2010, **62**, 391–403.
63 A. L. Schilmiller, G. D. Moghe, P. Fan, B. Ghosh, J. Ning, A. D. Jones and R. L. Last, *Plant Cell*, 2015, **27**, 1002–1017.
64 P. Fan, A. M. Miller, A. L. Schilmiller, X. Liu, I. Ofner, A. D. Jones, D. Zamir and R. L. Last, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, E239–E248.
65 J. Kim, K. Kang, E. Gonzales-Vigil, F. Shi, A. D. Jones, C. S. Barry and R. L. Last, *Plant Physiol.*, 2012, **160**, 1854–1870.
66 J. B. Landis, C. M. Miller, A. K. Broz, A. A. Bennett, N. Carrasquilla-Garcia, D. R. Cook, R. L. Last, P. A. Bedinger and G. D. Moghe, *Mol. Biol. Evol.*, 2021, **38**, 3202–3219.
67 P. Fan, A. M. Miller, X. Liu, A. D. Jones and R. L. Last, *Nat. Commun.*, 2017, **8**, 2080.
68 B. J. Leong, D. B. Lybrand, Y.-R. Lou, P. Fan, A. L. Schilmiller and R. L. Last, *Sci. Adv.*, 2019, **5**, eaaw3754.
69 J. Ning, G. D. Moghe, B. Leong, J. Kim, I. Ofner, Z. Wang, C. Adams, A. D. Jones, D. Zamir and R. L. Last, *Plant Physiol.*, 2015, **169**, 1821–1835.
70 P. Fan, P. Wang, Y.-R. Lou, B. J. Leong, B. M. Moore, C. A. Schenck, R. Combs, P. Cao, F. Brandizzi, S.-H. Shiu and R. L. Last, *eLife*, 2020, **9**, e56717.
71 B. J. Leong, S. M. Hurney, P. D. Fiesel, G. D. Moghe, A. D. Jones and R. L. Last, *Plant Physiol.*, 2020, **183**, 915–924.
72 Y.-R. Lou, T. M. Anthony, P. D. Fiesel, R. E. Arking, E. M. Christensen, A. D. Jones and R. L. Last, *Sci. Adv.*, 2021, **7**, eabj8726.
73 O. T. Chortyk, S. J. Kays and Q. Teng, *J. Agric. Food Chem.*, 1997, **45**, 270–275.
74 O. T. Chortyk, R. F. Severson, H. C. Cutler and V. A. Sisson, *Biosci., Biotechnol., Biochem.*, 1993, **57**, 1355–1356.
75 R. R. King and L. A. Calhoun, *Phytochemistry*, 1988, **27**, 3761–3763.
76 T. Matsuzaki, Y. Shinozaki, S. Suhara, M. Ninomiya, H. Shigematsu and A. Koiwai, *Agric. Biol. Chem.*, 1989, **53**, 3079–3082.
77 Y. Herrera-Salgado, M. L. Garduño-Ramírez, L. Vázquez, M. Y. Rios and L. Alvarez, *J. Nat. Prod.*, 2005, **68**, 1031–1036.
78 S. S. Nadakuduti, J. B. Uebler, X. Liu, A. D. Jones and C. S. Barry, *Plant Physiol.*, 2017, **175**, 36–50.
79 F. Zhou and E. Pichersky, *Curr. Opin. Plant Biol.*, 2020, **55**, 1–10.
80 L. Pazouki and U. Niinemets, *Front. Plant Sci.*, 2016, **7**, 1019.
81 P. S. Karunanithi and P. Zerbe, *Front. Plant Sci.*, 2019, **10**, 1166.
82 F. Zhou and E. Pichersky, *New Phytol.*, 2020, **226**, 1341–1360.
83 T. A. Akhtar, Y. Matsuba, I. Schauvinhold, G. Yu, H. A. Lees, S. E. Klein and E. Pichersky, *The Plant Journal*, 2013, **73**, 640–652.
84 A. M. Boutanaev, T. Moses, J. Zi, D. R. Nelson, S. T. Mugford, R. J. Peters and A. Osbourn, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, E81–E88.
85 Y. Matsuba, T. T. H. Nguyen, K. Wiegert, V. Falara, E. Gonzales-Vigil, B. Leong, P. Schaefer, D. Kudrna, R. A. Wing, A. M. Bolger, B. Usadel, A. Tissier, A. R. Fernie, C. S. Barry and E. Pichersky, *Plant Cell*, 2013, **25**, 2022–2036.
86 J. Zi, Y. Matsuba, Y. J. Hong, A. J. Jackson, D. J. Tantillo, E. Pichersky and R. J. Peters, *J. Am. Chem. Soc.*, 2014, **136**, 16951–16953.
87 E. Gonzales-Vigil, D. E. Hufnagel, J. Kim, R. L. Last and C. S. Barry, *The Plant Journal*, 2012, **71**, 921–935.
88 A. L. Schilmiller, I. Schauvinhold, M. Larson, R. Xu, A. L. Charbonneau, A. Schmidt, C. Wilkerson, R. L. Last and E. Pichersky, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 10865–10870.
89 C. Sallaud, D. Rontein, S. Onillon, F. Jabès, P. Duffé, C. Giacalone, S. Thoraval, C. Escoffier, G. Herbette, N. Leonhardt, M. Causse and A. Tissier, *Plant Cell*, 2009, **21**, 301–317.
90 J.-H. Kang, E. Gonzales-Vigil, Y. Matsuba, E. Pichersky and C. S. Barry, *Plant Physiol.*, 2014, **164**, 80–91.
91 P. M. Bleeker, P. J. Diergaarde, K. Ament, S. Schütz, B. Johne, J. Dijkink, H. Hiemstra, R. de Gelder, M. T. J. de

Published on 24 March 2022. Downloaded by Michigan State University on 10/17/2023 4:02:57 PM.

Both, M. W. Sabelis, M. A. Haring and R. C. Schuurink, *Phytochemistry*, 2011, **72**, 68–73.
92 P. M. Bleeker, R. Mirabella, P. J. Diergaarde, A. VanDoorn, A. Tissier, M. R. Kant, M. Prins, M. de Vos, M. A. Haring and R. C. Schuurink, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 20124–20129.
93 R. M. Coates, J. F. Denissen, J. A. Juvik and B. A. Babka, *J. Org. Chem.*, 1988, **53**, 2186–2192.
94 J. E. Frelichowski and J. A. Juvik, *J. Econ. Entomol.*, 2001, **94**, 1249–1259.
95 S. Zabel, W. Brandt, A. Porzel, B. Athmer, S. Bennewitz, P. Schäfer, R. Kortbeek, P. Bleeker and A. Tissier, *Plant Journal*, 2021, **105**, 1309–1325.
96 H. Duranova, V. Valkova and L. Gabriny, *Phytochem. Rev.*, 2021, DOI: 10.1007/s11101-021-09789-7.
97 J. R. Friedman, S. D. Richbart, J. C. Merritt, K. C. Brown, K. L. Denning, M. T. Tirona, M. A. Valentovic, S. L. Miles and P. Dasgupta, *Biomed. Pharmacother.*, 2019, **118**, 109317.
98 F. Spiller, M. K. Alves, S. M. Vieira, T. A. Carvalho, C. E. Leite, A. Lunardelli, J. A. Poloni, F. Q. Cunha and J. R. de Oliveira, *J. Pharm. Pharmacol.*, 2008, **60**, 473–478.
99 S. Varghese, P. Kubatka, L. Rodrigo, K. Gazdikova, M. Caprnda, J. Fedotova, A. Zulli, P. Kruzliak and D. Büsselberg, *Int. J. Food Sci. Nutr.*, 2017, **68**, 392–401.
100 J. J. Tewksbury and G. P. Nabhan, *Nature*, 2001, **412**, 403–404.
101 M. J. Caterina, A. Leffler, A. B. Malmberg, W. J. Martin, J. Trafton, K. R. Petersen-Zeitz, M. Koltzenburg, A. I. Basbaum and D. Julius, *Science*, 2000, **288**, 306–313.
102 S. E. Jordt and D. Julius, *Cell*, 2002, **108**, 421–430.
103 C. C. García, M. H. J. Barfuss, E. M. Sehr, G. E. Barboza, R. Samuel, E. A. Moscone and F. Ehrendorfer, *Ann. Bot.*, 2016, **118**, 35–51.
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. 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.-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. 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. 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.-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. Kim, H.-S. Pai, H. K. Ahn, I. Yeam, J. J. Giovannoni, J. K. C. Rose, I. Sørensen, S.-J. Lee, R. W. Kim, I.-Y. Choi, B.-S. Choi, J.-S. Lim, Y.-H. Lee and D. Choi, *Nat. Genet.*, 2014, **46**, 270–278.
105 P. Tripodi, M. T. Rabanus-Wallace, L. Barchi, S. Kale, S. Esposito, A. Acquadro, R. Schafleitner, M. van Zonneveld, J. Prohens, M. J. Diez, A. Börner, J. Salinier, B. Caromel, A. Bovy, F. Boyaci, G. Pasev, R. Brandt, A. Himmelbach, E. Portis, R. Finkers, S. Lanteri, I. Paran, V. Lefebvre, G. Giuliano and N. Stein, *Proc. Natl. Acad. Sci. U. S. A.*, 2021, **118**, e2104315118.
106 K. Han, H. Y. Lee, N. Y. Ro, O. S. Hur, J. H. Lee, J. K. Kwon and B. C. Kang, *Plant Biotechnol. J.*, 2018, **16**, 1546–1558.
107 M. Park, J. H. Lee, K. Han, S. Jang, J. Han, J. H. Lim, J. W. Jung and B. C. Kang, *Theor. Appl. Genet.*, 2019, **132**, 515–529.
108 Y. Q. Lang, H. Kisaka, R. Sugiyama, K. Nomura, A. Morita, T. Watanabe, Y. Tanaka, S. Yazawa and T. Miwa, *The Plant Journal*, 2009, **59**, 953–961.
109 Y. Tanaka, T. Sonoyama, Y. Muraga, S. Koeda, T. Goto, Y. Yoshida and K. Yasuba, *Mol. Breed.*, 2015, **35**, 142.
110 N. Weber, A. Ismail, M. Gorwa-Grauslund and M. Carlquist, *BMC Biotechnol.*, 2014, **14**, 25.
111 S. Koeda, K. Sato, H. Saito, A. J. Nagano, M. Yasugi, H. Kudoh and Y. Tanaka, *Theor. Appl. Genet.*, 2019, **132**, 65–80.
112 C. Stewart, B. C. Kang, K. Liu, M. Mazourek, S. L. Moore, E. Y. Yoo, B. D. Kim, I. Paran and M. M. Jahn, *The Plant Journal*, 2005, **42**, 675–688.
113 J. G. Roddick, M. Weissenberg and A. L. Leonard, *Phytochemistry*, 2001, **56**, 603–610.
114 L. C. Dolan, R. A. Matulka and G. A. Burdock, *Toxins*, 2010, **2**, 2289–2332.
115 X.-Y. Gu, X.-F. Shen, L. Wang, Z.-W. Wu, F. Li, B. Chen, G.-L. Zhang and M.-K. Wang, *Phytochemistry*, 2018, **147**, 125–131.
116 Y. Iijima, B. Watanabe, R. Sasaki, M. Takenaka, H. Ono, N. Sakurai, N. Umemoto, H. Suzuki, D. Shibata and K. Aoki, *Phytochemistry*, 2013, **95**, 145–157.
117 P. D. Cárdenas, P. D. Sonawane, U. Heinig, A. Jozwiak, S. Panda, B. Abebie, Y. Kazachkova, M. Pliner, T. Unger, D. Wolf, I. Ofner, E. Vilaprinyo, S. Meir, O. Davydov, A. Gal-On, S. Burdman, A. Giri, D. Zamir, T. Scherf, J. Szymanski, I. Rogachev and A. Aharoni, *Nat. Commun.*, 2019, **10**, 5169.
118 J. R. Paudel, C. Davidson, J. Song, I. Maxim, A. Aharoni and H. H. Tai, *Mol. Plant-Microbe Interact.*, 2017, **30**, 876–885.
119 B. A. Shinde, B. B. Dholakia, K. Hussain, S. Panda, S. Meir, I. Rogachev, A. Aharoni, A. P. Giri and A. C. Kamble, *Plant Mol. Biol.*, 2017, **95**, 411–423.
120 M.-C. Sánchez-Mata, W. E. Yokoyama, Y.-J. Hong and J. Prohens, *J. Agric. Food Chem.*, 2010, **58**, 5502–5508.
121 P. D. Sonawane, J. Pollier, S. Panda, J. Szymanski, H. Massalha, M. Yona, T. Unger, S. Malitsky, P. Arendt, L. Pauwels, E. Almekias-Siegl, I. Rogachev, S. Meir, P. D. Cárdenas, A. Masri, M. Petrikov, H. Schaller, A. A. Schaffer, A. Kamble, A. P. Giri, A. Goossens and A. Aharoni, *Nat. Plants*, 2017, **3**, 16205.
122 L. Barchi, M. Pietrella, L. Venturini, A. Minio, L. Toppino, A. Acquadro, G. Andolfo, G. Aprea, C. Avanzato, L. Bassolino, C. Comino, A. Dal Molin, A. Ferrarini, L. C. Maor, E. Portis, S. Reyes-Chin-Wo, R. Rinaldi, T. Sala, D. Scaglione, P. Sonawane, P. Tononi, E. Almekias-Siegl, E. Zago, M. R. Ercolano, A. Aharoni, M. Delledonne, G. Giuliano, S. Lanteri and G. L. Rotino, *Sci. Rep.*, 2019, **9**, 11769.
123 P. D. Sonawane, U. Heinig, S. Panda, N. S. Gilboa, M. Yona, S. P. Kumar, N. Alkan, T. Unger, S. Bocobza, M. Pliner, S. Malitsky, M. Tkachev, S. Meir, I. Rogachev and

 This journal is © The Royal Society of Chemistry 2022

Published on 24 March 2022. Downloaded by Michigan State University on 10/17/2023 4:02:57 PM.

A. Aharoni, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, E5419–E5428.
124 R. Akiyama, H. J. Lee, M. Nakayasu, K. Osakabe, Y. Osakabe, N. Umemoto, K. Saito, T. Muranaka, Y. Sugimoto and M. Mizutani, *Plant Biotechnol.*, 2019, **36**, 253–263.
125 H. J. Lee, M. Nakayasu, R. Akiyama, M. Kobayashi, H. Miyachi, Y. Sugimoto, N. Umemoto, K. Saito, T. Muranaka and M. Mizutani, *Plant Cell Physiol.*, 2019, **60**, 1304–1315.
126 M. Nakayasu, R. Akiyama, M. Kobayashi, H. J. Lee, T. Kawasaki, B. Watanabe, S. Urakawa, J. Kato, Y. Sugimoto, Y. Iijima, K. Saito, T. Muranaka, N. Umemoto and M. Mizutani, *Plant Cell Physiol.*, 2020, **61**, 21–28.
127 J. Szymański, S. Bocobza, S. Panda, P. Sonawane, P. D. Cárdenas, J. Lashbrooke, A. Kamble, N. Shahaf, S. Meir, A. Bovy, J. Beekwilder, Y. Tikunov, I. Romero de la Fuente, D. Zamir, I. Rogachev and A. Aharoni, *Nat. Genet.*, 2020, **52**, 1111–1121.
128 R. Akiyama, M. Nakayasu, N. Umemoto, J. Kato, M. Kobayashi, H. J. Lee, Y. Sugimoto, Y. Iijima, K. Saito, T. Muranaka and M. Mizutani, *Plant Cell Physiol.*, 2021, **62**, 775–783.
129 Y. Kazachkova, I. Zemach, S. Panda, S. Bocobza, A. Vainer, I. Rogachev, Y. Dong, S. Ben-Dor, D. Veres, C. Kanstrup, S. K. Lambertz, C. Crocoll, Y. Hu, E. Shani, S. Michaeli, H. H. Nour-Eldin, D. Zamir and A. Aharoni, *Nat. Plants*, 2021, **7**, 468–480.
130 N. Kim, O. Estrada, B. Chavez Jr, C. Stewart and J. C. D'Auria, *Molecules*, 2016, **21**, 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, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 10304–10309.
133 A. Mato, J. Limeres, I. Tomás, M. Muñoz, C. Abuin, J. F. Feijoo and P. Diz, *Br. J. Clin. Pharmacol.*, 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, *J. Palliative Med.*, 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 Phytol.*, 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, *Chem. Biol.*, 2006, **13**, 513–520.
139 A. Junker, J. Fischer, Y. Sichhart, W. Brandt and B. Dräger, *Front. Plant Sci.*, 2013, **4**, 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 Physiol.*, 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, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, 6133–6138.
143 M. A. Bedewitz, A. D. Jones, J. C. D'Auria and C. S. Barry, *Nat. Commun.*, 2018, **9**, 5281.
144 J.-P. Huang, C. Fang, X. Ma, L. Wang, J. Yang, J. Luo, Y. Yan, Y. Zhang and S.-X. Huang, *Nat. Commun.*, 2019, **10**, 4036.
145 R. S. Nett, Y. Dho, Y.-Y. Low and E. S. Sattely, *Proc. Natl. Acad. Sci. U. S. A.*, 2021, **118**, e2102949118.
146 K. Nakajima, T. Hashimoto and Y. Yamada, *Proc. Natl. Acad. Sci. U. S. A.*, 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, *Org. Lett.*, 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, *Nat. Commun.*, 2013, **4**, 2833.
150 T. Hashimoto and Y. Yamada, *Plant Physiol.*, 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 Catal.*, 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 Physiol.*, 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, *Mol. Plant*, 2020, **13**, 1203–1218.
155 E. R. Naves, L. D. Silva, R. Sulpice, W. L. Araújo, A. Nunes-Nesi, L. E. P. Peres and A. Zsögön, *Trends Plant Sci.*, 2019, **24**, 109–120.
156 P. Christen, S. Cretton, M. Humam, S. Bieri, O. Muñoz and P. Joseph-Nathan, *Phytochem. Rev.*, 2020, **19**, 615–641.
157 T. P. Michael and R. VanBuren, *Curr. Opin. Plant Biol.*, 2020, **54**, 26–33.
158 J.-K. Weng, R. N. Philippe and J. P. Noel, *Science*, 2012, **336**, 1667–1670.
159 H. A. Bunzel, X. Garraboul, M. Pott and D. Hilvert, *Curr. Opin. Struct. Biol.*, 2018, **48**, 149–156.
160 J. R. Jacobowitz and J.-K. Weng, in *Annual Review of Plant Biology*, ed. S. S. Merchant, 2020, vol. 71, pp. 631–658.
161 A. Onoyovwe, J. M. Hagel, X. Chen, M. F. Khan, D. C. Schriemer and P. J. Facchini, *Plant Cell*, 2013, **25**, 4110–4122.
162 V. Courdavault, N. Papon, M. Clastre, N. Giglioli-Guivarc'h, B. St-Pierre and V. Burlat, *Curr. Opin. Plant Biol.*, 2014, **19**, 43–50.

Published on 24 March 2022. Downloaded by Michigan State University on 10/17/2023 4:02:57 PM.

163 K. H. Ryu, L. Huang, H. M. Kang and J. Schiefelbein, *Plant Physiol.*, 2019, **179**, 1444–1456.
164 C. Seyfferth, J. Renema, J. R. Wendrich, T. Eekhout, R. Seurinck, N. Vandamme, B. Blob, Y. Saeys, Y. Helariutta, K. D. Birnbaum and B. De Rybel, in *Annual Review of Plant Biology*, ed. S. S. Merchant, 2021, vol. 72, pp. 847–866.
165 B. M. Moore, P. Wang, P. Fan, B. Leong, C. A. Schenck, J. P. Lloyd, M. D. Lehti-Shiu, R. L. Last, E. Pichersky and S.-H. Shiu, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 2344–2353.
166 B. M. Moore, P. Wang, P. Fan, A. Lee, B. Leong, Y.-R. Lou, C. A. Schenck, K. Sugimoto, R. Last, M. D. Lehti-Shiu, C. S. Barry and S.-H. Shiu, *In silico Plants*, 2020, **2**, diaa005.
167 P. Wang, B. M. Moore, S. Uygun, M. D. Lehti-Shiu, C. S. Barry and S.-H. Shiu, *New Phytol.*, 2021, **231**, 475–489.
168 L. P. de Souza, S. Alseekh, F. Scossa and A. R. Fernie, *Nat. Methods*, 2021, **18**, 733–746.
169 R. L. Last, A. D. Jones and Y. Shachar-Hill, *Nat. Rev. Mol. Cell Biol.*, 2007, **8**, 167–174.
170 H. Horai, M. Arita, S. Kanaya, Y. Nihei, T. Ikeda, K. Suwa, Y. Ojima, K. Tanaka, S. Tanaka, K. Aoshima, Y. Oda, Y. Kakazu, M. Kusano, T. Tohge, F. Matsuda, Y. Sawada, M. Y. Hirai, H. Nakanishi, K. Ikeda, N. Akimoto, T. Maoka, H. Takahashi, T. Ara, N. Sakurai, H. Suzuki, D. Shibata, S. Neumann, T. Iida, K. Tanaka, K. Funatsu, F. Matsuura, T. Soga, R. Taguchi, K. Saito and T. Nishioka, *J. Mass Spectrom.*, 2010, **45**, 703–714.
171 H. Tsugawa, R. Nakabayashi, T. Mori, Y. Yamada, M. Takahashi, A. Rai, R. Sugiyama, H. Yamamoto, T. Nakaya, M. Yamazaki, R. Kooke, J. A. Bac-Molenaar, N. Oztolan-Erol, J. J. B. Keurentjes, M. Arita and K. Saito, *Nat. Methods*, 2019, **16**, 295–298.
172 M. Wang, J. J. Carver, V. V. Phelan, L. M. Sanchez, N. Garg, Y. Peng, N. Don Duy, J. Watrous, C. A. Kapono, T. Luzzatto-Knaan, C. Porto, A. Bouslimani, A. V. Melnik, M. J. Meehan, W. T. Liu, M. Criisemann, P. D. Boudreau, E. Esquenazi, M. Sandoval-Calderon, R. D. Kersten, L. A. Pace, R. A. Quinn, K. R. Duncan, C.-C. Hsu, D. J. Floros, R. G. Gavilan, K. Kleigrewe, T. Northen, R. J. Dutton, D. Parrot, E. E. Carlson, B. Aigle, C. F. Michelsen, L. Jelsbak, C. Sohlenkamp, P. Pevzner, A. Edlund, J. McLean, J. Piel, B. T. Murphy, L. Gerwick, C.-C. Liaw, Y.-L. Yang, H.-U. Humpf, M. Maansson, R. A. Keyzers, A. C. Sims, A. R. Johnson, A. M. Sidebottom, B. E. Sedio, A. Klitgaard, C. B. Larson, C. A. Boya P, D. Torres-Mendoza, D. J. Gonzalez, D. B. Silva, L. M. Marques, D. P. Demarque, E. Pociute, E. C. O'Neill, E. Briand, E. J. N. Helfrich, E. A. Granatosky, E. Glukhov, F. Ryffel, H. Houson, H. Mohimani, J. J. Kharbush, Y. Zeng, J. A. Vorholt, K. L. Kurita, P. Charusanti, K. L. McPhail, K. F. Nielsen, L. Vuong, M. Elfeki, M. F. Traxler, N. Engene, N. Koyama, O. B. Vining, 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, *Nat. Biotechnol.*, 2016, **34**, 828–837.
173 L. W. Sumner, Z. Lei, B. J. Nikolau and K. Saito, *Nat. Prod. Rep.*, 2015, **32**, 212–229.
174 Y. Dong, P. Sonawane, H. Cohen, G. Polturak, L. Feldberg, S. H. Avivi, I. Rogachev and A. Aharoni, *New Phytol.*, 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, Genet.*, 2013, **3**, 1003–1013.
177 P. Gramazio, J. Prohens, M. Plazas, G. Mangino, F. J. Herraiz and S. Vilanova, *Front. Plant Sci.*, 2017, **8**, 1477.
178 M. Sulli, L. Barchi, L. Toppino, G. Diretto, T. Sala, S. Lanteri, G. L. Rotino and G. Giuliano, *Front. Plant Sci.*, 2021, **12**, 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 Phytol.*, 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, *Nat. Commun.*, 2015, **6**, 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 Cent. Sci.*, 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, *Angew. Chem., Int. Ed.*, 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, *Nat. Chem. Biol.*, 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, *Metab. Eng.*, 2017, **42**, 185–193.

 This journal is © The Royal Society of Chemistry 2022