Running title: Trichome metabolic diversity

Trading acyls and swapping sugars: metabolic innovations in *Solanum* trichomes

Paul D. Fiesel[1], Rachel E. Kerwin[1], A. Daniel Jones[1], and Robert L. Last[1,2]

[1]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48823 USA

[2]Department of Plant Biology, Michigan State University, East Lansing, MI 48823 USA

Corresponding Author: Robert L. Last, Michigan State University, Department of Biochemistry and Molecular Biology, 603 Wilson Road, East Lansing, MI 48824; Phone: 517-432-3278; email: lastr@msu.edu

ORCID:

Paul D. Fiesel 0000-0002-2038-5417

Rachel E. Kerwin 0000-0001-7395-5973

A. Daniel Jones 0000-0002-7408-6690

Robert L. Last 0000-0001-6974-9587

Abstract

Solanaceae (nightshade family) species synthesize a remarkable array of clade- and tissue-specific specialized metabolites. Protective acylsugars, one such class of structurally diverse metabolites, are produced from sugars and acyl-Coenzyme A esters by acylsugar acyltransferases in glandular trichomes. We characterized trichome acylsugars of the Clade II species *Solanum melongena* (brinjal eggplant) using liquid chromatography-mass spectrometry (LC-MS), gas chromatography-MS and nuclear magnetic resonance (NMR) spectroscopy. This led to the identification of eight unusual structures with inositol cores, inositol glycoside cores, and hydroxyacyl chains. LC-MS analysis of 31 species in the megadiverse *Solanum* genus revealed striking acylsugar diversity with some traits restricted to specific clades and species. The Acylinositols were found throughout each clade while acylglucoses were restricted to DulMo and VANAns species. Medium length hydroxyacyl chains were identified in many species. Analysis of tissue-specific transcriptomes and interspecific acylsugar acetylation differences led to the unexpected identification of the *S. melongena* Acylsugar AcylTransferase 3-Like 1 (SmASAT3-L1; SMEL4.1_12g015780) enzyme. This enzyme is distinct from previously characterized acylsugar acetyltransferases, which are in the ASAT4 clade, and is a functionally divergent ASAT3. This study provides a foundation for investigating the evolution of diverse *Solanum* acylsugar structures and harnessing this diversity in breeding and synthetic biology.

Introduction

Plants are remarkable synthetic chemists, producing a multitude of structurally complex specialized metabolites that differ from the products of general, or primary, metabolism in their lineage-specific distribution and tissue or cell type-specific biosynthesis. In contrast to the negative selection against changes in primary metabolism (for example, amino acids, energy metabolism intermediates and vitamin cofactors), less constrained evolution of specialized metabolism led to accumulation of hundreds of thousands of taxonomically restricted metabolites in broad classes. Specialized metabolites play many critical roles such as in abiotic and biotic stress adaptation (Agati and Tattini, 2010; De Moraes et al., 2001; Landry et al., 1995), pollinator attraction (Kretschmar and Baumann, 1999) and mediation of interactions with beneficial and pathogenic microbes (Yu et al., 2021). These diverse and bioactive small molecules have historical and modern uses in human medicine, including the anticancer alkaloids vinblastine and paclitaxel, antimalarial artemisinin, and painkillers such as morphine.

Acylsugars are specialized metabolites produced across the Solanaceae (nightshade) family, aiding in defense against herbivores, fungi, and bacteria (Goffreda et al., 1989; Leckie et al., 2016; Luu et al., 2017; Weinhold and Baldwin, 2011). In Type I- and IV-glandular trichomes, BAHD-type AcylSugar AcylTransferase (ASAT) enzymes assemble acylsugars from the basic building blocks of sugars, often sucrose, and short- to medium-length acyl chains derived from acyl-CoAs (Fan et al., 2015; Lou et al., 2021; Moghe et al., 2017; Schilmiller et al., 2012; Schilmiller et al., 2015). Despite their simple components, acylsugars exhibit remarkable chemical diversity arising from variations in sugar core composition and acyl chain length, branching pattern, position, and number (Fan et al., 2019; Fiesel et al., 2022; Ghosh et al., 2014a;

Hurney, 2018; Lou et al., 2021; Moghe et al., 2017; Schenck et al., 2022). For example, acylsucroses, composed of a sucrose disaccharide core, accumulate in cultivated tomato (*Solanum lycopersicum*) trichomes, while acylsucroses and acylglucoses have been observed in the trichomes of wild tomato species, *Solanum pennellii*. Acylsugar structural variation also impacts biological activity; for example, differential oviposition deterrence was demonstrated from naturally derived acylsugar mixtures (Leckie et al., 2016).

Solanaceae acylsugars have become an exemplary model to study evolution of a diverse, biologically relevant trait in a plant family with genomic and phylogenetic resources. Utilization of this model revealed gene duplication, neofunctionalization, co-option, and loss involved in acylsugar evolution (Fan et al., 2020, 2017; Leong et al., 2019; Moghe et al., 2017). For example, neofunctionalization of an invertase-like enzyme, AcylSucrose FructoFuranosidase 1 (ASFF1) (Leong et al., 2019) and functional divergence of core ASAT enzymes is responsible for differences in sugar core type and acyl chain positions between cultivated tomato *S. lycopersicum* and wild tomato *S. pennellii* acylsugars (Fan et al., 2017). Identifying mechanisms of acylsugar evolution across the broader *Solanum* requires a detailed understanding of acylsugar diversity and biosynthesis, which is lacking for many species.

Nearly half of the Solanaceae falls into the large (>1200 species) and phenotypically diverse *Solanum* genus. This genus is split into several major clades, including Potato, Regmandra, DulMo, VANAns, and Clade II (Figure 5); the Potato clade contains cultivated tomato and potato and their wild relatives while Clade II contains the cultivated brinjal eggplant, *Solanum melongena*, and other ‘spiny *Solanums*’ (Bohs, 2004; Bohs and Olmstead, 1997; Gagnon et al.,

2022; Levin et al., 2006; PBI Solanum Project, 2022; Särkinen et al., 2013; Stern et al., 2011; Tepe et al., 2016; Weese and Bohs, 2007). To date, documentation of acylsugar diversity largely focused on a handful of species within the Potato clade, including cultivated tomato and its close relatives. These efforts identified at least 38 nuclear magnetic resonance (NMR) spectroscopy-resolved acylsugar structures and liquid chromatography mass spectrometry (LC-MS) supported annotations of many more (Fan et al., 2017; Ghosh et al., 2014a; Lybrand et al., 2020; Schilmiller et al., 2016). While limited acylsugar screening outside of the Potato clade was reported, novel structural variants not observed among cultivated tomato relatives were identified. For example, acylsugars with *myo*-inositol sugar cores (i.e., acylinositols) were characterized in three species: *S. nigrum*, from the DulMo clade, and *S. lanceolatum* and *S. quitoense*, from Clade II (Herrera-Salgado et al., 2005; Hurney, 2018; Leong et al., 2020; Lou et al., 2021). These discoveries highlight the benefits of a more comprehensive description of *Solanum* acylsugars within the well-developed phylogenetic framework.

Here we report analysis of *Solanum* acylsugar chemical diversity in species from the relatively unexplored *Solanum* clades DulMo, VANAns, and Clade II, which together comprise ~1000 *Solanum* species. We first established the Clade II brinjal eggplant, *S. melongena*, as a reference species. Eggplant is a major worldwide fruit crop with extensive genomic, transcriptomic, and germplasm resources (Barchi et al., 2021; Li et al., 2021; Mennella et al., 2010). We characterized eggplant trichome acylinositols, acylinositol glycosides, and acylsugars with unusual hydroxylated acyl chains using electrospray ionization LC-quadrupole time of flight-MS (ESI LC-QToF-MS) and NMR. These atypical structures likely reflect biochemistry that is differentiated from the cultivated tomato acylsucrose pathway. Moving beyond this model

organism framework, LC-MS phylogenetic screening of 31 Clade II, DulMo, and VANAns species, including *S. melongena*, led to the identification of remarkable acylsugar structural variation. LC-MS features with characteristics of acylinositols were found in 25 of the 26 acylsugar-producing species, suggesting one or a small number of evolutionary origins. In contrast, acylglucoses were detected in DulMo and VANAns species, but not in any tested Clade II species.

As a first step towards unraveling the molecular basis underlying the extensive documented acylsugar diversity, we utilized interspecific acylsugar differences and an eggplant tissue-specific transcriptome to identify an acylinositol biosynthetic enzyme, *S. melongena* Acylsugar AcylTransferase 3-Like 1 (SmASAT3-L1; SMEL4.1_12g015780), responsible for acetylating a triacylinositol glycoside acyl acceptor. SmASAT3-L1 exhibits a different acyl-CoA specificity than previously characterized ASAT3 homologs, highlighting how gene duplication and functional divergence created acylsugar metabolic novelty in this part of the *Solanum* clade.

Results and Discussion

Eggplant glandular trichomes accumulate acylsugar-like metabolites

We began our investigation of *Solanum* acylsugar diversity with the brinjal eggplant *S. melongena* due to its economic importance, genomic resources, and phylogenetic position within the monophyletic Eastern Hemisphere Spiny clade of Clade II (formerly known as the ‘Old World spiny clade’) (Gagnon et al., 2022). Using eight eggplant accessions (Table S37), we observed glandular trichomes that esemble the acylsugar-producing structures found in other *Solanum* species on hypocotyls, cotyledons, and the first three true leaves of young eggplant

plants (Figure 2) (Leong et al., 2020; Lou et al., 2021; Schilmiller et al., 2012). In contrast, we observed only non-glandular stellate trichomes on leaves and stems of mature eggplants (Figure S12), and these are unlikely to accumulate and synthesize acylsugars or other specialized metabolites (Levin, 1973; Wagner, 1991).

We analyzed surface metabolite extracts from young and mature eggplant tissues using LC-QToF-MS coupled with collision-induced dissociation (CID) in negative and positive ion mode, and annotated acylsugars based on relative mass defect, molecular adduct ion masses, retention times, and ions present in CID mass spectra. Briefly, acylsugars were annotated from masses of fatty acid carboxylate fragment ions, as well as fragment ions corresponding to stepwise losses of acyl chains from the pseudomolecular ion to a sugar core fragment ion. The acylsugar annotation methods and confidence criteria are explained in detail in the Methods. This analysis revealed abundant acylsugars in extracts from young, glandular trichome-producing eggplant tissues, but not from mature, non-glandular trichome producing tissues.

We annotated 38 acylsugars in young eggplant extracts from eight accessions, including 16 acylhexoses and 22 acyldisaccharides (Table 1). LC-MS based acylsugar annotations are described using a modified shorthand nomenclature (Leong et al., 2020) as follows: UX:Y:Z(A,B,C,D), where U – as a single or multi-letter designation – represents the sugar core, X represents the number of acyl chains, Y represents the sum of acyl chain carbons, Z represents the number of unsaturated bonds in the acyl chains (when present), and A-D represent the number of carbon atoms in the individual acyl chains. For example, the eggplant acylhexose I3:18(4,4,10) consists of a *myo*-inositol core with three acyl chains with a total of 18 carbon

atoms. Curiously, eggplant acyldisaccharides contain an atypical pentose-hexose core, as evidenced by a fragment ion mass of *m/z* 293.09 in negative-ion mode, corresponding to a fully deacylated sugar core minus a hydroxyl group and a proton. Further evidence was provided by positive mode CID, which promotes glycosidic bond cleavage, yielding a fragment ion corresponding to the neutral loss of an unacylated pentose ring. All acylhexoses annotated with medium to high confidence formed abundant fragment ions in positive mode but few fragment ions in negative mode as illustrated in Figure S1; this pattern is characteristic of acylinositols found in *S. quitoense* and *S. nigrum*, but has not been observed for acylglucoses, suggesting that all detected *S. melongena* acylhexoses are acylinositols (Hurney, 2018; Leong et al., 2020; Lou et al., 2021). The eggplant acylsugars contained three to four acylations, all on the hexose core, including one medium eight-carbon (C8) to C14 acyl chain and two to three short C4 or C5 acyl chains. The medium acyl chains (C8, C10, C12, and C14) were either iso-branched or straight as revealed by GC-MS acyl chain analysis (Figure S15). Additionally, we identified hydroxylated C12, C14, and C16 acyl chains not previously reported in Solanaceae acylsugars. Although we did not observe large differences between the eight eggplant accessions, eggplant acylsugars differ in chain length, functional groups, and acyl chain composition from the reported *Solanum* acylinositols (Herrera-Salgado et al., 2005; Hurney, 2018; Lou et al., 2021).

Table 1. Summary of annotated acylsugars detected in *S. melongena* leaf surface extracts. PH = pentose-hexose; AI = arabinose-inositol, I = inositol. RT = retention time; m/z_{acc} = theoretical monoisotopic formate adduct mass; m/z_{ex} = experimental formate adduct mass; Δm (ppm) = mass measurement error in parts per million. Compound number is listed for NMR characterized compounds (Figure 2). Percent total peak area calculated by dividing acylsugar peak area by total acylsugar peak area. Acylsugars composing ≥1% of total acylsugar peak area are bolded. Percent peak area calculated from 15 *S. melongena* leaf surface extracts. Annotation method is described in Materials and Methods. Acylsugars are sorted by number of sugar moieties and then by elution order.

Name	Compound #	Annotation confidence level	RT (min)	Neutral molecule chemical formula	m/z_{acc}	m/z_{ex}	Δm (ppm)	Average percent peak area (%)
Acyldisaccharides								
PH2:12(4,8)		medium	1.98	C23H40O12	553.2496	553.2496	0.0	0.203
PH4:14(2,4,4,4)		medium	2.19	C25H40O14	609.2395	609.2445	8.2	0.0246
PH3:14(2,4,8)		medium	2.23	C25H42O13	595.2602	595.2621	3.2	0.0453
PH4:16(2,2,4,8)		medium	2.54	C27H44O14	637.2708	637.2720	1.9	0.0851
PH3:16(4,4,8)		**medium**	**2.61**	**C27H46O13**	**623.2915**	**623.2916**	**0.2**	**17.1**
AI3:16(4,4,8)	**1**	**high**	**2.66**	**C27H46O13**				
PH4:17(2,2,4,9)		medium	2.93	C28H46O14	651.2864	651.2924	9.2	0.114
PH3:17(4,5,8)		medium	2.96	C28H48O13	637.3071	637.3083	1.9	0.324
PH4:18(2,4,4,8)		medium	3.11	C29H48O14	665.3021	665.3033	1.8	0.901
AI4:18(2,4,4,8)	**2**	**high**	**3.27**	**C29H48O14**	**665.3021**	**665.3026**	**0.7**	**42.5**
PH3:18(4,4,10)		**medium**	**3.61**	**C29H50O13**	**651.3228**	**651.3205**	**-3.5**	**1.12**
PH3:18(4,4,10)		medium	3.74	C29H50O13	651.3228	651.3245	2.5	0.964
PH4:19(2,4,5,8)		medium	3.76	C30H50O14	679.3177	679.3146	-4.6	0.329
PH4:19(2,4,5,8)		medium	3.83	C30H50O14				
PH4:20(4,4,4,8)		medium	4.22	C31H52O14	693.3334	693.3350	2.3	0.0428
PH4:20(2,4,4,10)		medium	4.57	C31H52O14	693.3334	693.3351	2.5	0.0447
PH4:20(2,4,4,10)		medium	4.68	C31H52O14	693.3334	693.3349	2.1	0.127
PH4:20(2,4,4,10)		medium	4.85	C31H52O14	693.3334	693.3379	6.6	0.00422
PH3:20(4,4,12)		medium	5.05	C31H54O13	679.3541	679.3554	1.9	0.0389
PH3:20(4,4,12)		medium	5.23	C31H54O13	679.3541	679.3552	1.6	0.130
PH3:22(4,4,14)		medium	7.54	C33H58O13	707.3854	707.3879	3.5	0.0356
PH3:22(4,4,14)		medium	7.85	C33H58O13	707.3854	707.3892	5.3	0.0107
Acylhexoses								
I3:16(4,4,8)		**medium**	**3.27**	**C22H38O9**	**491.2492**	**491.2492**	**-0.1**	**3.99**
I3:20(4,4,12-OH)		medium	4.16	C26H46O10	563.3068	563.3071	0.5	0.0808
I3:20(4,4,12-OH)		medium	4.32	C26H46O10	563.3068	563.3068	0.0	0.500
I4:18(2,4,4,8)		medium	4.22	C24H40O10	533.2598	533.2604	1.1	0.0248
I3:18(4,4,10)	**3**	**high**	**4.70**	**C24H42O9**	**519.2805**	**519.2802**	**-0.5**	**10.6**
I3:18(4,4,10)		medium	4.88	C24H42O9	519.2805	519.2794	-2.1	0.693
I3:20(2,4,14-OH)		medium	4.68	C26H46O10	563.3068	563.3082	2.5	0.0588
I3:20(2,4,14-OH)		medium	4.90	C26H46O10	563.3068	563.3077	1.6	0.298
I3:22(4,4,14-OH)	**7**	**high**	**6.06**	**C28H50O10**	**591.3381**	**591.3385**	**0.7**	**4.96**
I3:22(4,4,14-OH)	**8**	**high**	**6.34**	**C28H50O10**	**591.3381**	**591.3380**	**-0.1**	**10.6**
I3:20(4,4,12)	**4**	**high**	**6.81**	**C26H46O9**	**547.3118**	**547.3119**	**0.1**	**2.24**
I3:20(4,4,12)		medium	7.09	C26H46O9	547.3118	547.3127	1.7	0.105
I3:24(4,4,16-OH)		medium	8.51	C30H54O10	619.3694	619.3707	2.0	0.0473
I3:24(4,4,16-OH)		medium	8.85	C30H54O10	619.3694	619.3683	-1.8	0.0355
I3:22(4,4,14)	5	high	9.34	C28H50O9	575.3431	575.3435	0.7	0.986

I3:22(4,4,14)	6	high	9.68	C28H50O9	575.3431	575.3433	0.3	0.722

Although LC-MS provided information about sugar core mass, it did not reveal the sugar core structure, prompting analysis using gas chromatography-mass spectrometry (GC-MS) of derivatized *S. melongena* acylsugar cores. When free sugar cores, produced by metabolite extract saponification, were derivatized to form alditol acetates, GC-MS peaks corresponding to derivatized *myo*-inositol and glucose were detected, supporting the presence of acylinositols (Figure 1A). The detection of glucose might have resulted from other compounds in the leaf surface metabolite extracts. The disaccharide sugar core composition was determined by hydrolyzing the saponified sugar cores with formic acid to cleave the glycosidic linkage. Comparison of the hydrolyzed and unhydrolyzed samples revealed a peak corresponding to arabinose only in the hydrolyzed plant samples (Figure 1B). Taken together, the results of saponification with and without hydrolysis, followed by derivatization, confirmed identification of *myo*-inositol sugar cores, and identified the pentose moiety of the hexose-pentose disaccharide core as arabinose.

NMR analysis of eight eggplant acylsugars

While MS analysis provided valuable information about acyl chain number and length, sugar core mass, as well as acyl chain number, complementary information about acyl chain branch structure, sugar core stereochemistry, and acyl chain positions was obtained using NMR. We purified and resolved the structures of eight abundant eggplant acylsugars from accession PI 555598 using a combination of 1D and 2D NMR experiments (Figure 2). All eight structures are newly described, and because atomic connections were determined by NMR and MS data, the

proposed structures meet Metabolomics Standards Initiative level 1 criteria for metabolite identification (Sumner et al., 2007).

A series of NMR experiments confirmed that the acylhexose and acyldisaccharide sugar cores are *myo*-inositol and 4-*O*-β-arabinopyranosyl *myo*-inositol, respectively (Figure 2C,D). We assigned all sugar ring proton signals of each sugar core's spin system with total correlation spectroscopy (TOCSY). Inferences from correlation spectroscopy (COSY) data then identified the order of ring protons and revealed pyranose and cyclitol ring structures for the pentose and hexose rings, respectively. Relative stereochemistry at ring positions was subsequently determined through comparison of spin-spin splitting (multiplicities and coupling constants) referenced to expected patterns inferred from chemical principles and previously reported acylinositols (Hurney, 2018; Leong et al., 2020). The disaccharide glycosidic linkage position was determined by heteronuclear multiple bond correlation (HMBC) correlations to be at position 4 and position 1 of *myo*-inositol and arabinose, respectively. The arabinose *β* anomeric configuration was inferred from the anomeric carbon $^{1}J_{\mathrm{CH}}$ (162 Hz) as revealed by a coupled-heteronuclear single quantum coherence (coupled-HSQC) experiment of the free disaccharide sugar core after saponification. Our sugar core assignments identifying this unusual disaccharide agreed with the previous GC-MS sugar core results, supporting the efficacy of the sugar core GC-MS identification method. This disaccharide core differs in the pentose moiety identity from a recently reported acylated 4-*O*-*ß*-xylopyranosyl *myo*-inositol in the *Solanum* Clade II species *S. quitoense* (Hurney, 2018). To our knowledge, this is the first report of acylated 4-*O*-*ß*-arabinosyl *myo*-inositol sugars.

Acyl chain positions, branching patterns, and hydroxyl positions were also resolved through integration of different NMR experiments. We found that all acyl chains were confined to *myo*-inositol, consistent with the LC-MS results. All eight acylsugars are decorated with two short iso-branched iC4 acyl chain esters at positions 1 and 2, and one medium C8 to C14 acyl chain ester at position 3 (Figure 2). Compound 2 (Figure 2C), the only tetraacylated acylsugar identified in eggplant, additionally carried an acetylation at position 6 (Figure 2C). The medium-length acyl chains at position 3 were resolved as straight (nC14) or terminally iso-branched (iC8, iC10, iC12, iC14) based on signals characteristic of protons near the branched carbons. Strikingly, we identified peculiar hydroxylated straight and iso-branched 3-hydroxytetradecanoate acyl chains, 3-OH-nC14 and 3-OH-iC14, in compounds 7 and 8, respectively (Figure 2D). We assigned hydroxylation positions of 3-OH-nC14 and 3-OH-iC14 to the third acyl carbon based on a downfield shifted signal in the ^{1}H NMR spectrum at 3.92 ppm, corresponding to one hydrogen at that position. In contrast, the non-hydroxylated medium acyl chains observed in compounds 1-6 (Figure 2C,D), carry two hydrogen atoms at the third acyl carbon, and these have a characteristic signal near 1.50 ppm. We believe this is the first report of 3-OH-nC14 and 3-OH-iC14 chains in Solanaceae acylsugars. While 3-OH-nC14 chains are observed in the acylsugar-like bacterial Lipid A glycolipids, the hydroxylation position differentiates these eggplant chains from hydroxyacyl chains in castor bean (*Ricinus communis*) seed oil, *Silene gallica* gallicasides, *Ibicella lutea* fatty acid glycosides, and Convolvulaceae resin glycosides (Asai et al., 2010; Asai and Fujimoto, 2010; Bah and Pereda-Miranda, 1996; Smith, 1971).

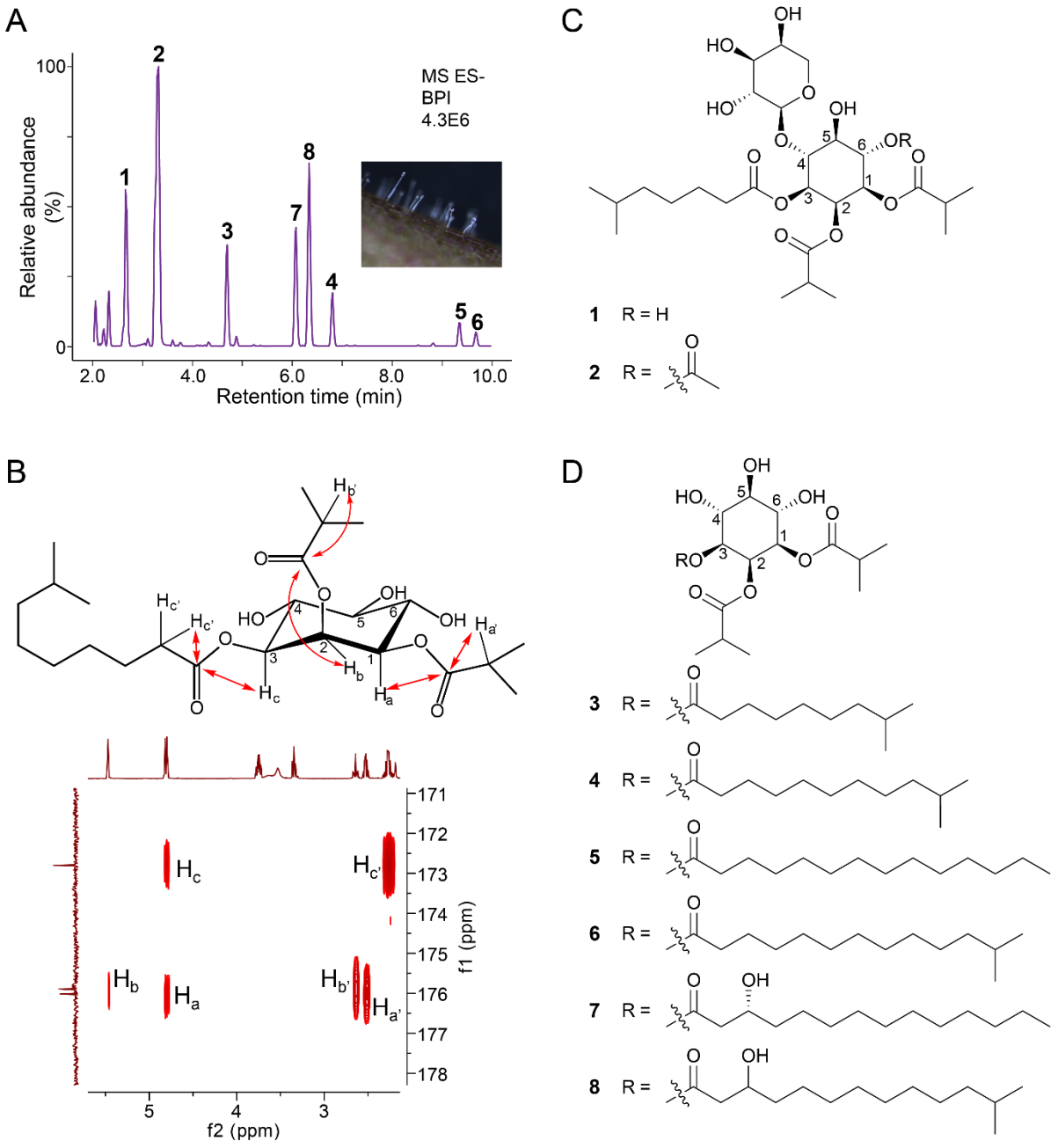

Figure 2. Profiling of *S. melongena* acylsugars. (A) LC-MS base peak intensity (BPI) chromatogram of *S. melongena* acylsugars. Peaks characterized by NMR are indicated with their compound numbers from Panels C and D. Metabolites were collected from seedling tissue with glandular trichomes similar to the ones displayed in the closeup photo of a *S. melongena* hypocotyl. (B) Annotation of HMBC experiments used to identify acylation positions of specific acyl chains. At the top, I3:18(4,4,10) is displayed in the chair conformation along with hydrogens important for determining acylation position. HMBC NMR spectrum showing

couplings between hydrogens and carbonyl carbons are shown. (C) NMR-characterized acyldisaccharide structures 1 and 2, with inositol ring carbons numbered. (D) NMR-characterized acylinositol monosaccharide structures 3-8.

The OH-tetradecanoate acyl chain has *R* stereochemistry

To gain insight into the biosynthetic origin of hydroxylated acyl chains, we tested chirality of the 3-OH-nC14 acyl chain of compound 7 as this chain has available commercial standards. Two-step chiral derivatization and GC-MS analysis was performed on compound 7 and two commercial standards, (3*R*)-OH-nC14 and the corresponding racemate (3*R*/*S*)-OH-nC14. Fatty acids were converted to ethyl esters to aid volatilization, followed by esterification of the acyl chain hydroxyl group with the Mosher acid (*R*)-(−)-α-methoxy-α-(trifluoromethyl)phenylacetate (*R*-MPTA) to distinguish stereoisomers by GC separation of the diastereomeric derivatives. GC-MS analysis of the derivatized (3*R*/*S*)-OH-nC14 standard compounds yielded two peaks at retention times of 42.21 and 42.40 minutes, while one peak at 42.40 minutes was obtained for (3*R*)-OH-nC14. We observed expected fragment ions corresponding to the MPTA fragment ion, fatty acid ethyl ester, and fatty acid which supported peak assignment (Figure S5). Thus (3*S*)-OH-nC14 and (3*R*)-OH-nC14 elute at 42.21 and 42.40 minutes, respectively. The compound 7 acyl chain was assigned as (3*R*)-OH-nC14 as its derivatized 3-OH-tetradecanoate acyl chain eluted at 42.40 minutes.

The (3*R*)-OH-nC14 chain hydroxyl position and stereochemistry have implications for the metabolic pathway leading to its biosynthesis. As fatty acid hydroxylases typically act on the acyl chain terminal region, they are unlikely to catalyze C3 hydroxylation of a C14 fatty acid (Pinot and Beisson, 2011). Of the other well-characterized pathways, only mitochondrial and plastidial fatty acid metabolism mechanisms produce (3*R*)-OH-acyl-thioester intermediates.

Consistent with this result, previous work identified two trichome-expressed, mitochondrial enzymes, Acylsugar Enoyl-CoA Hydratase 1 (AECH1) and Acylsugar Acyl-CoA Synthetase 1 (AACS1), and one trichome-expressed, plastidial enzyme, a beta-ketoacyl-(acyl-carrier-protein) reductase (SpKAR1), involved in medium C10 and C12 acyl chain production in cultivated tomato and its wild relative, *S. pennellii* (Fan et al., 2020; Ji et al., 2022). We hypothesize that acylsugar hydroxyacyl chains evolved through one of three scenarios. First, changes in acyl-CoA substrate availability, possibly through the actions of thioesterases or acyl-CoA synthetases, enabled a promiscuous ASAT to esterify hydroxyacyl chains to acylinositols. Alternatively, substrate specificity of an existing acylinositol ASAT expanded to utilize available hydroxyacyl-CoAs. Similar to this second hypothesis, amino acid substitutions between copies of ASAT2 from different tomato species impacted the accumulation of differently branched five carbon acyl chains on acylsugars (Fan et al., 2015). A third possibility is that a 'new' acyltransferase, unrelated to characterized ASATs and capable of utilizing hydroxyacyl-CoA substrates, was co-opted into acylinositol biosynthesis. Future genetic and biochemical studies should reveal the mechanisms behind (3*R*)-OH acyl chain production, potentially providing new approaches to engineer unique acylsugars with hydroxyacyl chains and hydroxylated fatty acids for polymeric building blocks. Acylsugar structures impact their function, thus we hypothesize that engineered tomato producing hydroxyacyl chains on their acylsugars may exhibit different defense

capabilities than the wildtype acylsucroses.

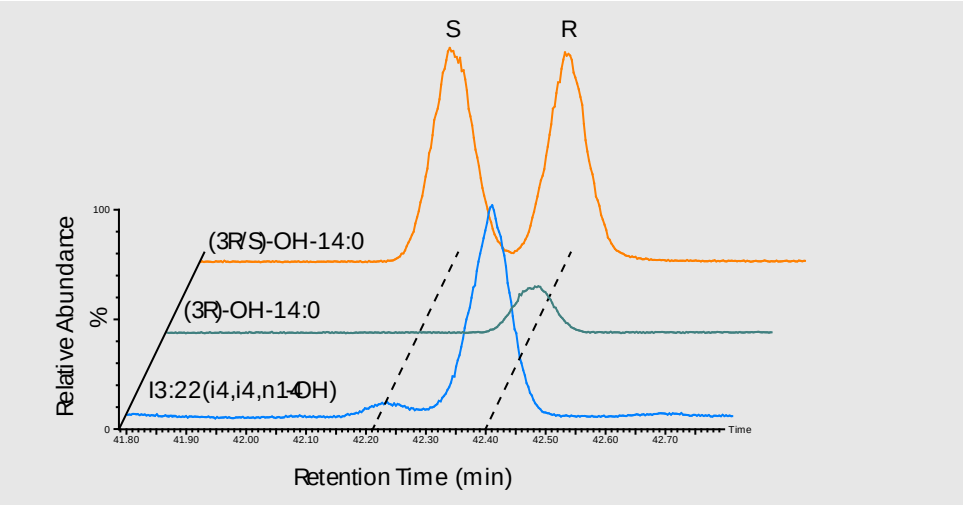

Figure 3. The 3-OH-n14 acyl chain on Compound 7, I3:22(i4,i4,n14-OH), is in the *R* configuration as revealed by GC-MS using a derivatization method that separates hydroxyacyl chain enantiomers. Hydroxyacyl chains analyzed by GC-selected ion monitoring-MS (GC-SIM-MS) after ethyl transesterification and derivation to their diastereomeric (*S*)-MPTA form. Fragment ions with *m/z* of 189, 209, and 255 were monitored, and relative abundance was normalized to the highest peak. Yellow and green traces represent derivatives from commercial (3*R*/*S*)- and (3*R*)-OH-14:0 standards. The main derivatized 3-OH-14:0 acyl chain from Compound 7 (Figure 2), shown with the blue trace, comigrates with the peak corresponding to the *R* enantiomer.

Enormous *Solanum* acylsugar diversity revealed by LC-MS metabolite screening

To date, most acylsugar screening across the *Solanum* has focused on cultivated tomato and its Potato clade relatives (Fan et al., 2017; Ghosh et al., 2014a; Lybrand et al., 2020; Schilmiller et al., 2016), which represent a small fraction of the species diversity in this genus (PBI Solanum Project, 2022). Though limited in number, studies profiling non-Potato clade species, including *S. nigrum* (Lou et al., 2021), *S. lanceolatum* (Herrera-Salgado et al., 2005), *S. quitoense* (Hurney, 2018; Leong et al., 2020), and *S. melongena* (this work) demonstrate acylsugar diversity across the *Solanum* is vast, and largely uncharacterized. To address this knowledge gap, we analyzed tissue surface extracts from 30 additional *Solanum* species, including 23 Clade II, five DulMo

clade, and two VANAns clade members (listed in Table S37). While NMR is the gold standard for structural elucidation, it is prohibitively time-consuming for a largescale metabolite diversity survey. LC-MS coupled with CID provides substantial structural information and is compatible with high-throughput screening. To strike a balance between data quality and quantity, we performed LC-MS-CID screening on all 30 species and annotated acylsugars as described for *S. melongena*. Twenty-five of the 30 analyzed species produced detectable acylsugars from visible surface glandular trichomes, and species lacking acylsugars had no or few observable glandular trichomes in the tissues analyzed. Targeted analysis of leaf surface extracts from 31 species, including *S. melongena*, and fruit surface extracts from two species, uncovered previously unreported acylsugars, inter-species acylsugar differences, and identified the distributions of unusual acylsugar chemical traits in the *Solanum* genus (Figure 5; Tables 1,S1-26). Although we did not observe intraspecific acylsugar variation, acylsugars varied enormously between species (Figure 4; Tables 1,S1-26).

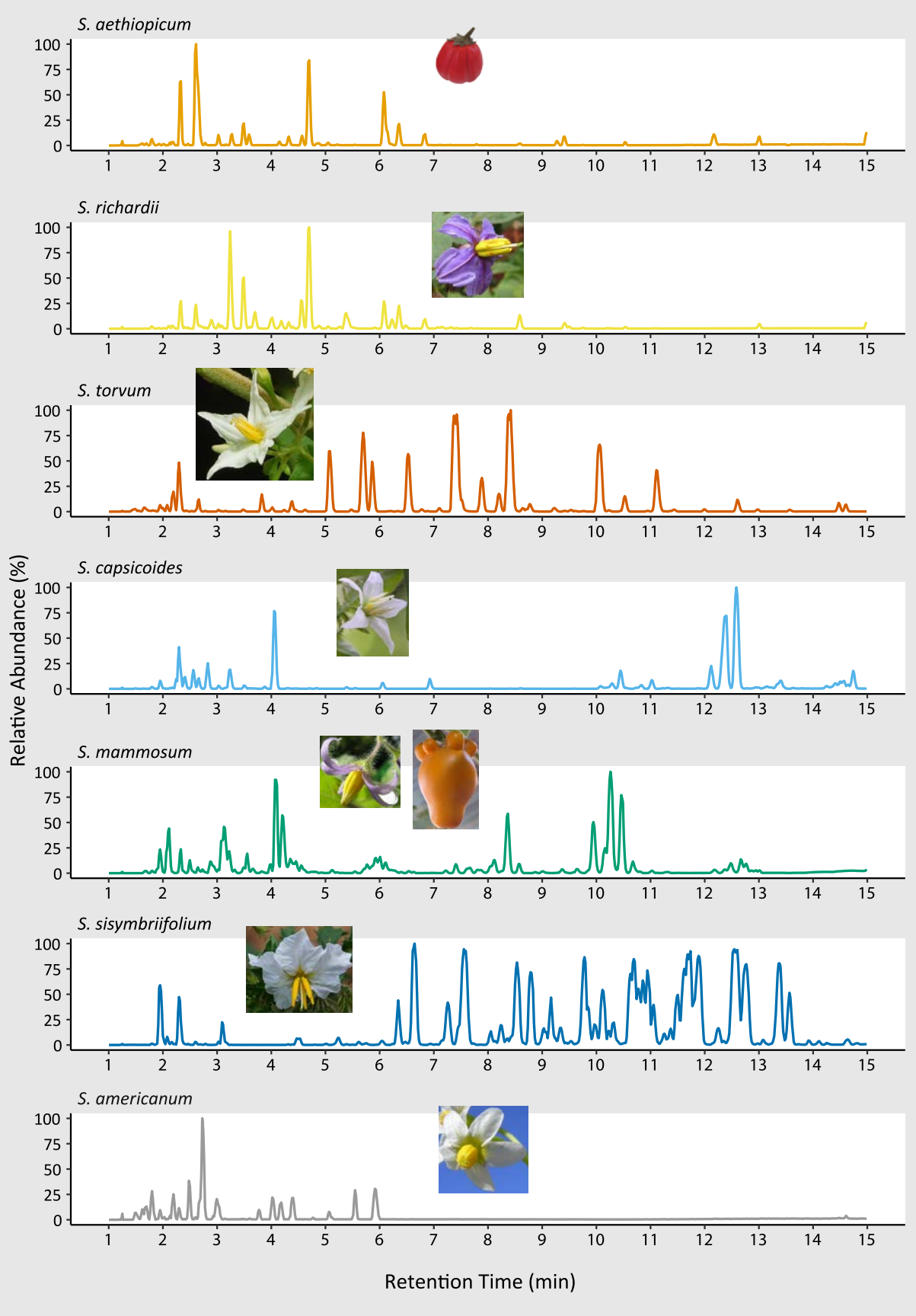
S. aethiopicum
S. richardii
S. torvum
S. capsicoides
S. mammosum
S. sisymbriifolium
S. americanum
Relative Abundance (%)
Retention Time (min)
0
25
50
75
100
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15

Figure 4. *Solanum* species produce dramatically different leaf surface metabolite profiles. The seven species display morphological differences, as demonstrated by the flower and fruit images, and they also exhibit diverse metabolite profiles as demonstrated by base peak intensity LC-MS chromatograms from their leaf surface metabolites. Most of the peaks shown consist of acylsugars with varied sugar cores and acyl chain types and numbers. Acylsugar profiles for these species are detailed in Tables S4, S7, S17, S20, and S24-26.

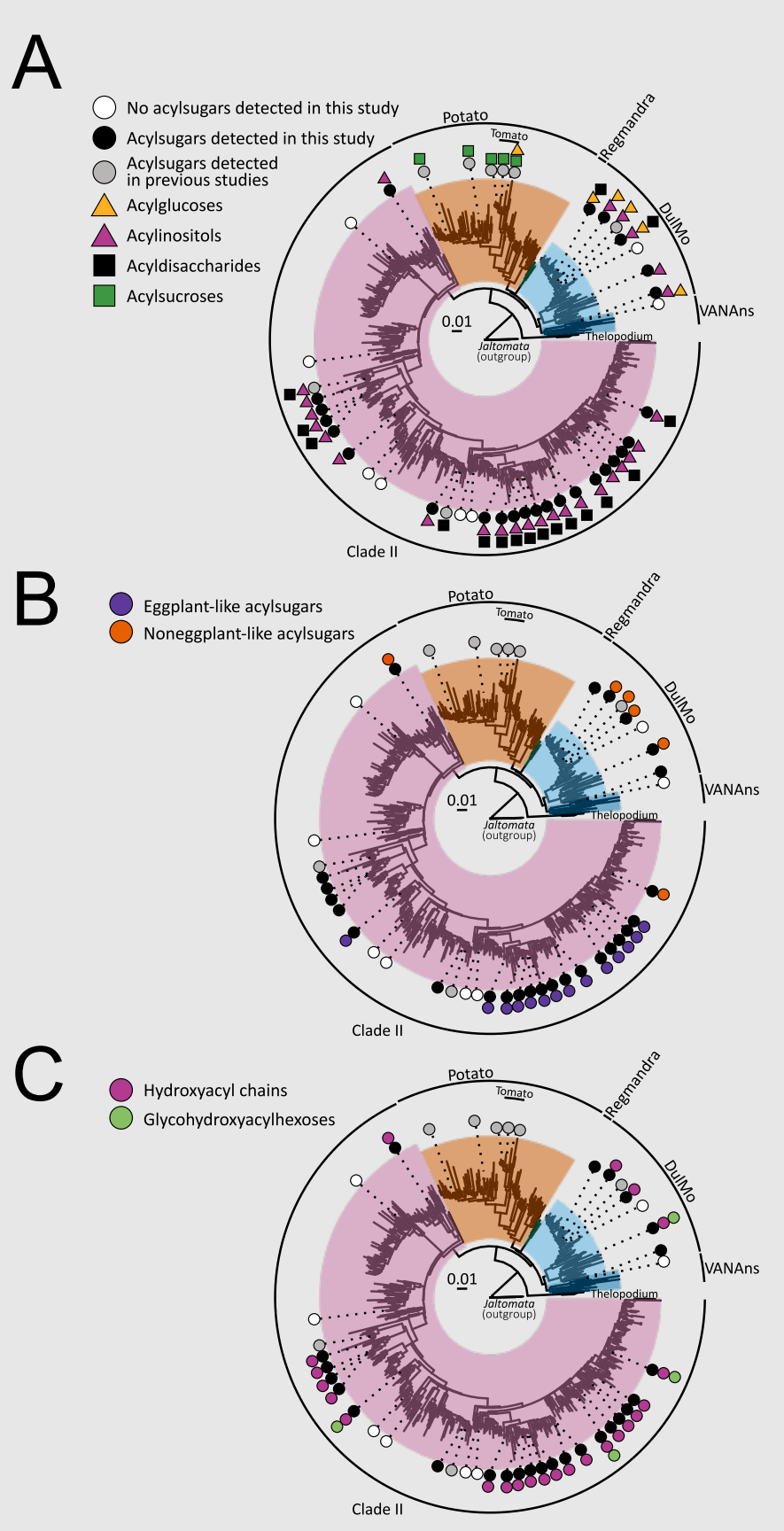
A
No acylsugars detected in this study
Acylsugars detected in this study
Acylsugars detected in previous studies
Acylglucoses
Acylinositols
Acyldisaccharides
Acylsucroses
Potato
Tomato
Regmandra
DulMo
VANAns
Thelopodium
0.01
Jaltomata
(outgroup)
Clade II
B
Eggplant-like acylsugars
Noneggplant-like acylsugars
Potato
Tomato
Regmandra
DulMo
VANAns
Thelopodium
0.01
Jaltomata
(outgroup)
Clade II
C
Hydroxyacyl chains
Glycohydroxyacylhexoses
Potato
Tomato
Regmandra
DulMo
VANAns
Thelopodium
0.01
Jaltomata
(outgroup)
Clade II

Figure 5. Distribution of acylsugar traits within the *Solanum* genus. Traits were mapped onto a previously published Supermatrix maximum likelihood phylogeny (Gagnon et al., 2022). The reported polytomy between Clade II, DulMO, VANANs, Regmandra, and Potato clades is not indicated (Gagnon et al., 2022). Acylsugar trait data from this report included annotations based upon LC-MS and NMR analysis (Figure 2, Tables 1, S2-26), while data from previous studies were limited to those with NMR structural data (Ghosh et al., 2014b; Herrera-Salgado et al., 2005; Hurney, 2018; King et al., 1986; Leong et al., 2020, 2019; Lou et al., 2021; Lybrand et al., 2020). (A) Distribution of sugar core types within the *Solanum* genus. Inositol and glucose cores determined by characteristic LC-MS negative mode electrospray ionization (ESI) fragmentation (Figures S1 and S7). (B) Distribution of species with acylsugars that coelute with eggplant acylsugars containing the same acyl chains and sugar core. (C) Distribution of hydroxyacyl chains and glycosylated hydroxyacyl chains.

Identification of diverse acyl chain types in *Solanum* acylsugars

Our survey of *Solanum* species uncovered acylsugars with a surprising diversity of acyl chain lengths, functional groups, and combinations of chain types and positions. We identified C2 to C18 acyl chains, including unsaturated and hydroxylated forms, based on analysis of fatty acid fragment carboxylate ions in negative mode and neutral losses in both negative and positive mode (Table S1). Acyl chain compositions differed between acylsugars within and between species: acylinositols with only medium C8 to C14 acyl chains were observed in 5 species (Figure 5, Tables S13-15, S20, S22, S24), while acylsugars with two short C4 to C6 acyl chains and one medium C8 to C14 acyl chain were observed in 17 species, an acylation pattern also observed in eggplant and wild tomato (Figures 2; Table 1) (Lybrand et al., 2020). Acylsugars with unsaturated acyl chains, including C5:1, C18:1, C18:2, and C18:3, were only detected in *S. torvum* (Table S17). To the best of our knowledge, this is the first report of unsaturated C18 chains in Solanaceae acylsugars. In contrast, hydroxylated medium-length acyl chains were surprisingly common: we detected acylsugars bearing hydroxylated acyl chains in three DulMo clade species and 25 Clade II species (Figure 5C). The additional hydroxyl group present on

these hydroxylated acyl chains is susceptible to further modifications, including glycosylations and acylations, which would result in even greater acylsugar diversity.

Indeed, we obtained evidence for glycosylated acyl chains in three Clade II species, *S. prinophyllum*, *S. sisymbriifolium*, *S. lasiophyllum*, and one DulMo species, *S. dulcamara*. Several lines of evidence led us to conclude that these species accumulate acylsugars with glycosyl groups attached via ether linkages to medium hydroxyacyl chains esterified to the hexose core. First, the compounds fragmented abnormally under negative mode CID: the glycosyl group was observed as a neutral loss from the $[M-H]^-$ ion, producing a $[sugar – H]^-$ fragment ion (Figure S2). This contrasts with extensive observations that CID spectra of other characterized acyldisaccharides do not exhibit negative mode disaccharide glycosidic linkage cleavage (Figures S1 and S8) (Ghosh et al., 2014; Hurney, 2018; Lou et al., 2021). Further evidence that the glycosylation is not on the primary acylated hexose ring was obtained by subjecting acylsugar extracts to saponification, which induces ester linkage breakage while retaining ether linkages. Upon analyzing saponified acylsugar extracts by LC-MS, we detected pentose and hexose sugars with C14 or C16 acyl chains rather than glycosylated hexoses, indicating that the glycosyl group is connected to a medium hydroxyacyl chain by an ether linkage that is not cleaved during alkaline saponification (Figure S3). We named these compounds glycohydroxyacylhexoses and annotated them by modifying the conventional acylsugar nomenclature where the molecule in Figure S2 is named H3:22(4,4,14-O-p) with 14-O-p representing the pentosylated (p) C14 hydroxyacyl chain (14-O). To our knowledge, this is the first report of acylhexoses with glycosylated hydroxyacyl chains in Solanaceae acylsugars. Resin glycosides produced by Convolvulaceae family species contain similar hydroxyacyl chain

glycosidic linkages. However, in resin glycosides, hydroxyacyl chains are connected to the oligosaccharide core at two points – by an ester and ether linkage, respectively – forming a macrocyclic structure not observed in the *Solanum* glycohydroxyacylhexoses (Bah and Pereda-Miranda, 1996; Kruse et al., 2022; Pereda-Miranda et al., 1993).

The unusual glycohydroxyacylhexoses raises questions regarding their biosynthetic and evolutionary origins. As these compounds are found in different lineages (Eastern Hemisphere Spiny, Dulcamaroid, and Sisymbriifolium sections) – and closely related species lack detectable glycohydroxyacylhexoses – this trait appears to be highly labile (Figure 5C). Because we observe the cognate, non-glycosylated acylinositol in each of these species, we hypothesize that glycohydroxyacylhexoses consist of a *myo*-inositol hexose core. Further analysis of acylsugars from additional species may identify other types of modified acylsugar hydroxyacyl chains in *Solanum* acylsugars.

Sixteen species accumulate acylsugars with the same molecular masses, sugar core masses, and acyl chain complements as the NMR-characterized eggplant acylinositols, suggesting they may be identical. To test this hypothesis, we mixed purified eggplant NMR-resolved standards with metabolite extracts from each species and performed liquid co-chromatography. Coelution provided one line of evidence that the eggplant-like acylsugars detected in 11 Clade II species are identical to characterized eggplant acylinositols (Tables S4-13,S26). In contrast, the eggplant-like acylsugars from five other species, including four DulMo clade members and one Clade II member, eluted separately from characterized eggplant acylinositols, indicating they are acylation positional isomers or acyl chain branching isomers (Figure S4). Examples of

acylinositol positional isomers were described between the DulMo clade species *S. nigrum* and the Clade II species *S. quitoense* and *S. melongena*, providing precedent that the non-coeluting compounds might be positional isomers of the eggplant acylsugars (Figure 2) (Leong et al., 2020; Lou et al., 2021). Our data reveal a phylogenetic pattern of acylinositol distribution in which Clade II species generally accumulate eggplant-like acylinositol isomers, whereas DulMo clade species accumulate noneggplant-like acylinositol isomers (Figure 5B). While the eggplant-like isomers coelute with eggplant acylsugars further analysis with complementary analytical methods is needed to confirm whether these are structurally identical across species.

Acylhexoses and acyldisaccharides accumulate throughout Clade II, DulMo, and VANAns *Solanum* species

Acylhexoses were common, being detected in all analyzed acylsugar-producing species. They were annotated as acylglucoses or acylinositols based upon differences in characteristic fragmentation behavior under negative ion mode MS conditions (Leong et al., 2020; Lou et al., 2021). Identification of acylglucoses was based on observation of a sugar core fragment ion of *m*/*z* 143.03 ($C_6H_7O_4^-$) and stepwise acyl chain losses at the lowest collision energy level (0 V) (Figure S7). In contrast, acylinositol annotation was based on an absence of acyl chain fragment ions at 0 V and MS^e CID negative mode functions (Figure S1). Using these rules, we identified acylinositols in all acylsugar-producing species tested except for *S. americanum* (Tables 1, S1-26). In contrast, acylglucoses were not detected in any Clade II species but were commonly detected among the DulMo and VANAns species, found in all analyzed acylsugar-producing species except *S. dulcamara* (Tables S1, S18, S19, S22, S23, S25; Figure 5A; Lou et al., 2021).

Differing from the widespread acylhexose distribution, acyldisaccharides were detected sporadically in Clade II and DulMo clade species (Figure 5A), being observed in 14 of 21 Clade II and 2 of 4 DulMo acylsugar-producing analyzed species. Based on negative $[M-H]^-$ fragment masses and positive mode fragmentation of the glycosidic linkage, the acyldisaccharides were composed of hexose-hexose, pentose-hexose, or deoxyhexose-hexose sugar cores (Table S1, Figures S1 and S8). Though complete structural information cannot be obtained from MS fragmentation alone, two lines of evidence lead us to hypothesize that these acyldisaccharides are composed of glycosylated inositol cores. First, most of the species with detectable acyldisaccharides also accumulate cognate acylinositols with the same acyl chain lengths (leaf and fruit surface extracts from 10 of 15 and two of two different species, respectively), suggesting a shared inositol-containing core structure (Tables S1, S2-8, S11-16, S19, S21, S25). Second, acylinositol glycosides with varying sugar core sizes and stereochemistries were described in *S. lanceolatum* (Herrera-Salgado et al., 2005), *S. quitoense* (Hurney, 2018), and *S. melongena* (Figure 2), suggesting that the species in this study may accumulate similar compounds. The ratio of acylhexose to acyldisaccharide peak number and abundance differs across the surveyed species; for example, *S. mammosum* only accumulated detectable acylinositols while *S. anguivi* primarily accumulated acyldisaccharides that comprised 80% of the total acylsugar peak area (Tables S20 and S3). This is reminiscent of varied acylhexose accumulation observed in trichomes of evolutionarily distant *Solanum* species: *S. nigrum* only contains detectable acylinositol and acylglucose monosaccharides (Lou et al., 2021), while *S. pennellii* accessions have mixtures of acylglucoses and acylsucroses ranging from 41-95% of total acylsugars as acylglucoses. (Lybrand et al., 2020; Shapiro et al., 1994).

Sticky fruits accumulate acylinositols

We investigated the chemical basis of sticky fruit surface substances previously reported in botanical species descriptions of the Clade II species *S. acerifolium* and *S. atropurpureum* (Nee, 2022a, 2022b, 1991). LC-MS analysis of fruit surface extracts revealed similar acylsugar profiles between the two species, with each extract containing what appeared to be the same 128 acylsugars. These fruit acylsugars were distinct from the 22 and 19 trichome acylsugars identified in *S. acerifolium* and *S. atropurpureum*, respectively, as evidenced by their LC elution and MS characteristics (Tables S13-16; Figure 6). Seventy-seven fruit acylsugars were identified as acylinositols with two or three medium C8 to C14 acyl chains, including one to three hydroxyacyl chains. The remaining 51 fruit acylsugars are acyldisaccharides containing an unusual deoxyhexose-hexose core not previously reported in Solanaceae acylsugars. In contrast, *S. acerifolium* and *S. atropurpureum* trichome acylsugars, composed of a pentose-hexose disaccharide core decorated with short C5 or C6 acyl chains, were not detected in fruit surface extracts.

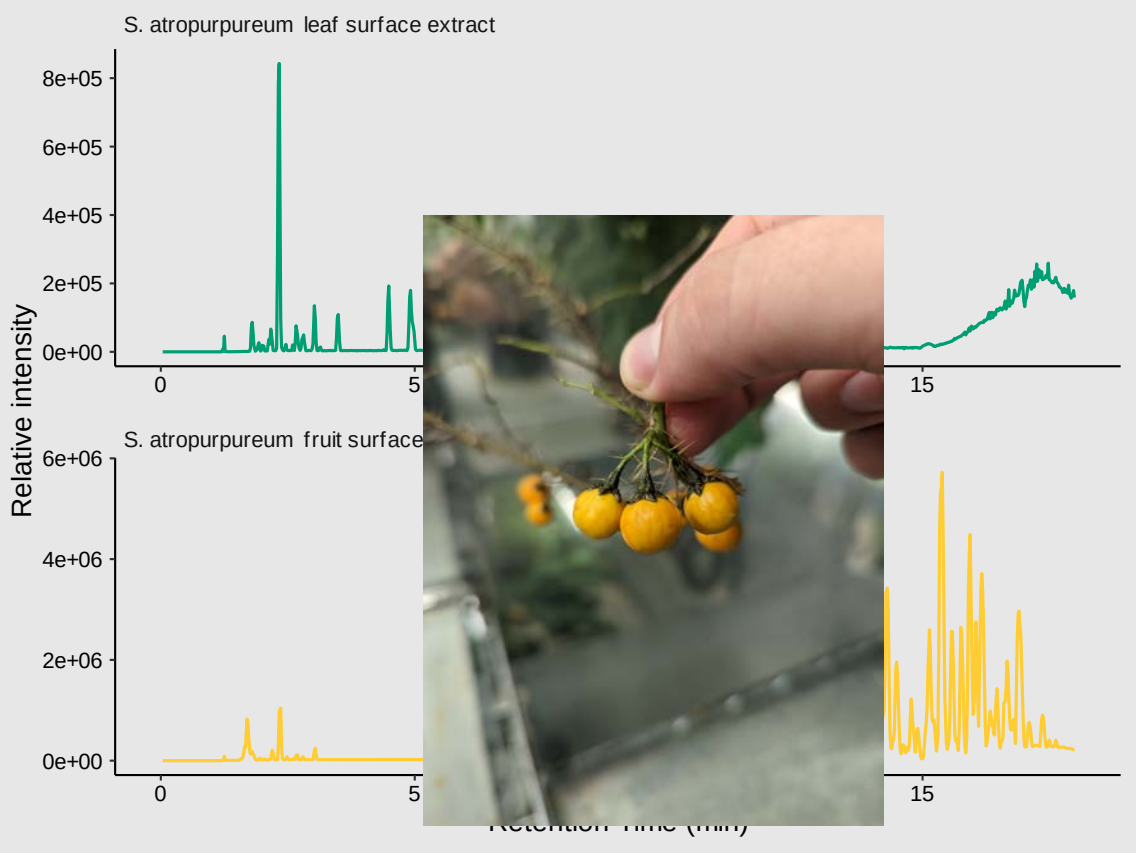

Figure 6. *S. atropurpureum* trichomes and fruit exhibit dramatically different metabolite profiles. The top LC-MS base peak intensity (BPI) chromatogram displays metabolites from a leaf surface extract, while the bottom LC-MS base peak intensity chromatogram displays metabolites from a fruit surface extract. The leaf surface extract acylsugars elute from 2.2 to 7.6 min (Table S16), while fruit surface extract acylsugars elute from 4.6 to 18.4 min (Table S14).

S. mammosum and *S. capsicoides*, which are closely related to *S. acerifolium* and *S. atropurpureum*, do not produce sticky fruit surfaces, suggesting this trait evolved in the common ancestor of *S. acerifolium* and *S. atropurpureum*, perhaps as recently as 2 Mya (Särkinen et al., 2013). Fruit surface acylsugars were described previously within the *Physalis* genus and likely represent an independent evolutionary origin from the *S. acerifolium* and *S. atropurpureum* fruit acylsugars (Bernal et al., 2018; Cao et al., 2015; Cicchetti et al., 2018; Maldonado et al., 2006). Considering that, like eggplant, *S. acerifolium* and *S. atropurpureum* fruits lack trichomes (Nee, 2022a, 2022b), some other organ or cell type presumably synthesizes the fruit surface acylsugars.

Structures similar to *Solanum fernadesii* petiolar resin glands (Sampaio et al., 2021), *Hypericum androsaemum* microscopic fruit glands (Caprioli et al., 2016), and fennel and chamomile fruit secretory ducts (vittae) (Zizovic et al., 2007) may be involved in *S. acerifolium* and *S. atropurpureum* fruit acylsugar secretion. Future work will be needed to understand the cellular, biosynthetic and evolutionary relationships between trichome and fruit acylsugars.

Sugar core evolution across the Solanaceae

Our screening revealed acylinositols to be broadly distributed across Clade II, DulMo, and VANAns (Figure 5A, Table S1). Outside of this study, acylinositols were only reported in *Solanum* species, including two Clade II members, *S. quitoense* and *S. lanceolatum*, and one DulMo clade member, *S. nigrum*. The acylinositol distribution suggests that acylinositol biosynthesis arose one or more times within the *Solanum* genus. However, our ability to elucidate the likely number of acylinositol origins is impeded by the lack of resolution of the phylogenetic relationships between the major *Solanum* clades, including Clade II, DulMo, VANAns, Potato, and Regmandra (Gagnon et al., 2022). Further biochemical and genetic analyses of acylinositol biosynthesis may reveal how this trait evolved, and whether the pathway is conserved between Clade II, DulMo, and VANAns species.

Acylglucoses were reported to be sporadically present across the Solanaceae, including in *Salpiglossis sinuata*, *Petunia*, *Nicotiana* spp., *Datura*, the DulMo clade *S. nigrum*, and Potato clade *S. pennellii* (Castillo et al., 1989; Chortyk et al., 1997; Fiesel et al., 2022; Fobes et al., 1985; Hurney, 2018; King and Calhoun, 1988; Lou et al., 2021; Matsuzaki et al., 1989; Schenck et al., 2022; Van Dam and Hare, 1998). Recent work in *S. pennellii* and *S. nigrum* revealed that

acylglucoses are synthesized from acylsucroses by a neofunctionalized invertase-like enzyme, AcylSucrose FructoFuranosidase 1 (ASFF1) (Leong et al., 2019; Lou et al., 2021). These results show that conversion of acylsucroses to acylglucoses arose at least two times in *Solanum* because *S. nigrum* and *S. pennellii* employ non-orthologous ASFF1 enzymes. Our current study extends documented cases of *Solanum* acylglucose distribution to the VANAns clade, as well as documenting it in additional DulMo species. In contrast, we found no acylglucose accumulating species among the 23 analyzed *Solanum* Clade II species. Acylglucose pathway determination of non-*Solanum* clade species will be needed to resolve the evolutionary history of these metabolites. Elucidation of acylglucose biosynthesis outside of the *Solanum* clade will be needed to resolve the number of evolutionary origins and biochemical mechanisms leading to synthesis of these metabolites across the Solanaceae.

Varied acyldisaccharides were observed across DulMo and Clade II, and we predict they are biosynthesized from acylinositols. Within the Potato clade as well as outside the *Solanum* genus, acylsucroses are the dominant acyldisaccharide in the reported species and tissues (Fiesel et al., 2022). In contrast, no acylsucroses were detected in any species screened. However, acylsucrose biosynthesis likely persists in DulMo and VANAns species considering that acylglucoses were detected in all but one DulMo and VANAns species and that acylsucroses can be intermediates in acylglucose biosynthesis (Leong et al., 2019; Lou et al., 2021). The lack of both acylglucoses and acylsucroses in the analyzed Clade II species suggests a loss of acylsucrose biosynthesis in Clade II. Rather, our LC-MS data are consistent with the hypothesis that the acyldisaccharides observed in this study consist of glycosylated inositol cores. Recent *in vitro* biochemistry and *in*

vivo genetic evidence in *S. quitoense* suggests that acylinositols are the precursors to acylinositol disaccharides in Clade II *Solanum* species (Leong et al., 2022, 2019). In this scenario, acylinositols would be converted to their cognate glycosides by one or more glycosyltransferases, with different sugars added either by a conserved promiscuous glycosyltransferase or multiple glycosyltransferases. *S. quitoense* and *S. melongena*, with their acylinositol glycosides bearing different glycosyl moieties, are promising systems for addressing the biochemical origins of these unusual *Solanum* acyldisaccharides.

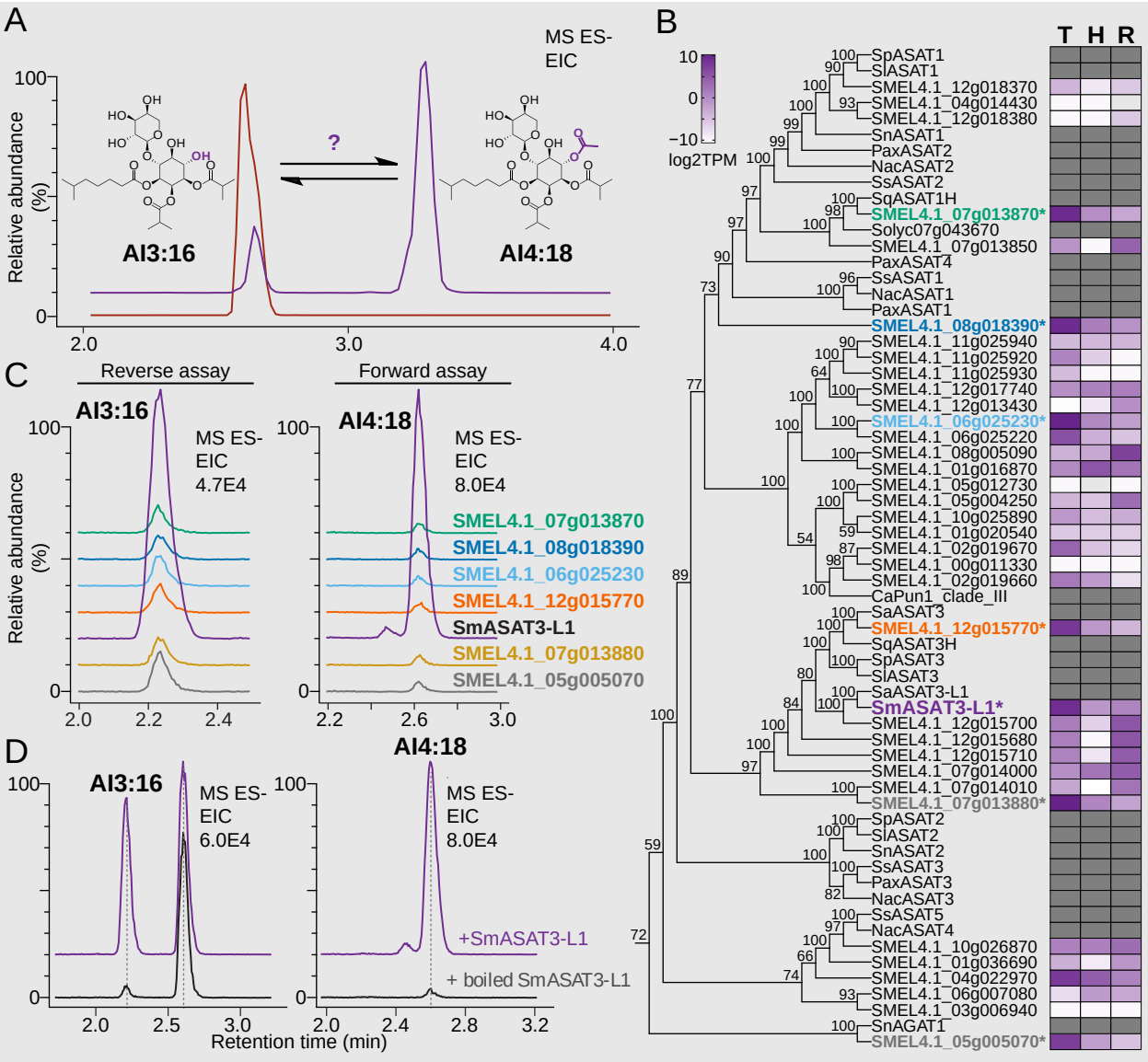

Figure 7. SmASAT3-L1 acetylates AI3:16 to produce AI4:18. (A) LC-MS analysis of AI3:16(4,4,8) and AI4(2,4,4,8) in *S. aethiopicum* (red) and *S. melongena* (purple) leaf surface extracts. *S. aethiopicum* does not produce detectable levels of AI4:18. (B) *S. melongena* Clade III BAHDs. Shown is a clade III BAHD tree subset from a phylogeny including 106 predicted BAHDs (PF002458) in the eggplant Smel_V4.1 reference genome, published reference BAHD sequences for clades I-VII, characterized ASAT sequences from other Solanaceae species, and the SaASAT3 and SaASAT3-L1 candidates from *S. aethiopicum* (see Figure S11 for full phylogeny). The maximum likelihood tree was inferred from amino acid sequences using the Jones-Taylor-Thornton algorithm with seven rate categories in IQ-TREE v2.1.3. Values at nodes indicate bootstrap support calculated from 100,000 ultrafast bootstrap replicates. Heatmap shows absolute transcript abundance (log2 TPM) across trichomes, trichomeless hypocotyls, and roots from 7-day-old eggplant seedlings. Expression data for non-eggplant sequences not included in heatmap. Color gradient provides a visual marker to rank the transcript abundance from high (purple) to low (white) or absent (grey). ASAT, acylsugar acyltransferase. log2TPM, log2 transformed transcripts per million. (C) Of the seven tested enzymes, only SmASAT3-L1 acetylates AI3:16(4,4,8) to form AI4:18(2,4,4,8). The extracted ion chromatograms on the left display products from forward enzyme assays and the formate adduct of AI4:18, *m/z* 665.30. The extracted ion chromatograms on the right display products from reverse enzyme assays and the formate adduct of AI3:16, *m/z* 623.29. (D) SmASAT3-L1 acetylates AI3:16 *in vitro* to produce AI4:18. (A) SmASAT3-L1 reaction scheme indicating acetylation position. (B) SmASAT3-L1 catalyzes the forward and reverse direction. Forward assay extracted ion chromatograms display the formate adduct of AI4:18, *m/z* 665.30, and reverse assay extracted ion chromatograms display the formate adduct of AI3:16, *m/z* 623.29. The reverse assay chromatograms display peaks for both AI3:16 and AI4:18 due to in source fragmentation of remaining AI4:18.

S. aethiopicum is defective in expression of a trichome acylsugar acyltransferase enzyme

We began analyzing the biochemical mechanisms underlying the observed *Solanum* acylsugar diversity in seven of 23 members from the Eggplant clade and related Anguivi grade, two small subclades within Clade II that includes brinjal eggplant and its close relatives (Figure S6) (Aubriot et al., 2018). In contrast to the other six analyzed eggplant species, *S. aethiopicum* (scarlet or Ethiopian eggplant) is unique in that it had detectable unacetylated AI3:16(i4,i4,i8) (Compound 1) but did not accumulate the acetylated form, AI4:18(2,i4,i4,i8) – the highest abundance *S. melongena* acylsugar (Compound 2; Figure S6, Tables S1, S3, S4). We hypothesized that the absence of AI4:18(2,i4,i4,i8) in *S. aethiopicum* is due either to loss-of-function or altered expression of an acetylating enzyme in this species. The availability of *S.*

melongena and *S. aethiopicum* genomic sequences (Barchi et al., 2021; Song et al., 2019) provided the opportunity to seek a trichome expressed acetyltransferase responsible for producing AI4:18 and the mechanism behind the lack of detectable AI4:18 in *S. aethiopicum*.

To identify enzyme candidates, we sequenced transcriptomes of isolated trichomes, trichome-depleted shaved hypocotyls, and whole roots collected from 7-day-old *S. melongena* seedlings and performed differential gene expression analysis. Of the 23,251 eggplant genes expressed in at least one of the three tissues, 745 were significantly enriched (log2 fold-change >2, *p*-value < 0.05) in trichomes compared to hypocotyls and roots, including 20 BAHDs (Supplemental datafile 1). We selected for further testing seven BAHDs that were abundantly expressed in trichomes (TPM > 200) and homologous to characterized *Solanum* ASATs (Figure 7B) (D'Auria, 2006; Fan et al., 2015; Leong et al., 2022, 2020; Lou et al., 2021; Schenck et al., 2022). We expressed each of the seven candidates in *Escherichia coli* and tested for acetylation of Compound 1, AI3:16(4,4,8). An SlASAT3 outparalog (Koonin, 2005) SMEL4.1_12g015780, ACYLSUGAR ACYLTRANSFERASE3-LIKE1 (SmASAT3-L1) was the only enzyme to exhibit forward activity acetylating Compound 1, AI3:16(4,4,8), to Compound 2, AI4:18(2,4,4,8) (Figure 7C). As shown in Figure S14, LC-MS analysis of the SmASAT3-L1 *in vitro* assay product revealed that it had identical molecular masses and elution times to plant-derived AI4:18(2,i4,i4,i8) (Compound 1) providing evidence that SmASAT3-L1 acetylates AI3:16(i4,i4,i8). Characterization of reverse activities of the seven BAHD acyltransferases (Leong et al., 2020; Lou et al., 2021), in which an acyl chain can be removed from an acylsugar by incubating it with the enzyme and free Coenzyme A, confirmed that SmASAT3-L1 was the only tested enzyme that removed an acetyl chain from AI4:18(2,4,4,8) (Figures 7D and S14).

Taken together, these results indicate that trichome-expressed *SmASAT3-L1* encodes an acylinositol acetyltransferase responsible for AI4:18(2,4,4,8) biosynthesis.

We tested the hypothesis that *S. aethiopicum* trichome extracts lack detectable acetylated AI4:18(2,4,4,8) due to a defect in *SmASAT3-L1* ortholog expression. Indeed, reverse transcription (RT)-PCR comparing isolated trichome and trichomeless shaved hypocotyl cDNA revealed that *SaASAT3-L1* (GenBank ID: OQ547782) transcript abundance was undetectable in *S. aethiopicum* trichomes (Figures S10). This result implicates the defect in ASAT3-L1 enzyme gene expression as responsible for lack of AI4:18 in *S. aethiopicum*. The interspecific variation in ASAT3-L1 expression is reminiscent of the ASAT4 expression differences among accessions of the wild tomato *Solanum habrochaites*. In this case, a subgroup of *S. habrochaites* accessions possessed a functional ASAT4 copy, but its low levels of gene expression correlated with reduced accumulation of acetylated acylsugars in these 'subgroup E' accessions (Kim et al., 2012).

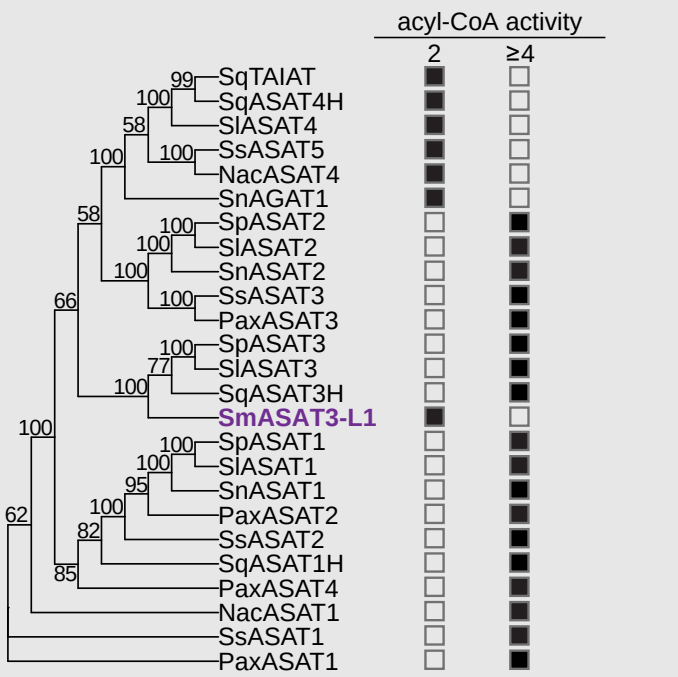

Figure 8. Characterized *in vitro* acyl-CoA usage by ASATs. The maximum likelihood phylogeny was inferred from amino acid sequences using a time-reversible algorithm specifying a plant-specific empirical substitution matrix, invariant sites, and four rate categories (Q.plant+I+G4) in IQ-TREE v2.1.3. Bootstrap support was calculated from 100,000 ultrafast bootstrap replicates; values >=95 indicate strong support. Other than SmASAT3-L1, enzyme activities were described previously (Fan et al., 2017, 2015; Leong et al., 2022, 2020; Lou et al., 2021; Moghe et al., 2017; Nadakuduti et al., 2017; Schenck et al., 2022; Schilmiller et al., 2015; Schilmiller et al., 2012). Only positive activities are displayed, and the blank squares may not indicate a lack of activity.

Natural variation potentiates novel activities and uses

Until recently, it appeared that acylsucroses dominate acylsugar profiles of Solanaceae species with a small number of documented examples of acylglucose- and acylinositol-producing *Solanum* species (Fiesel et al., 2022; Fobes et al., 1985; Herrera-Salgado et al., 2005; Hurney, 2018; King and Calhoun, 1988; Leong et al., 2020; Liu et al., 2017; Lou et al., 2021; Moghe et al., 2017; Schenck et al., 2022). Our chemotaxonomic survey of more than two dozen *Solanum* species from the sparsely sampled Clade II, DulMo, and VANAns clades revealed widespread occurrence of glucose-, inositol-, and non-sucrose disaccharide-based acylsugars decorated with unusual acyl chains, including medium-length (C12-C16) hydroxyacyl chains, glycosylated hydroxyacyl chains, and unsaturated chains. Considering that this work and previously published analyses together cover <3% of *Solanum* species, this is only the 'tip of the iceberg' of trichome acylsugar diversity within this large genus. Adding to interspecific variation we observe here, the recent identification of acylsugars in cultivated tomato root exudates supports the value of metabolite screening in additional tissue types as well as species (Korenblum et al., 2020).

Knowledge of *Solanum* acylsugar diversity provides a framework for investigating the molecular basis of metabolic pathway evolution. For example, the widespread occurrence of acylglucoses across the DulMo clade leads us to hypothesize that the neofunctionalized invertase responsible for *S. nigrum* acylglucose synthesis (Lou et al., 2021) is shared by other DulMo clade species. Similarly, documented acylsugar variation between cultivated eggplant, *S. melongena,* and its relative, *S. aethiopicum*, led to the discovery of a terminal acylsugar acetyltransferase in eggplant, SmASAT3-L1. The characteristics of ASAT3-L1, the enzyme absent from the *S. aethiopicum* trichome transcriptome, in turn reveals the dynamic nature of plant specialized metabolism. First, SmASAT3-L1 is unique among characterized ASATs in its acyl acceptor

specificity, acetylating an acylinositol glycoside rather than an acylated sucrose, glucose, or *myo*-inositol. Second, acetyl-CoA donor activity was previously only described for a subset of Clade III BAHDs, which include ASAT4 and ASAT5 clade members (Figure 8). SmASAT3-L1 breaks a pattern seen across the Solanaceae, including trichome ASATs from species of early evolving lineages, *Petunia axillaris* and *Salpiglossis sinuata* (Moghe et al., 2017; Nadakuduti et al., 2017), *Nicotiana attenuata* (Schenck et al., 2022), as well as cultivated and wild tomatoes (Kim et al., 2012; Schilmiller et al., 2012). This theme was also observed in trichomes of two other *Solanum* plants, more closely related to the species screened in this study: the Clade II *S. quitoense* SqTAIAT (Leong et al., 2020) and DulMo *S. nigrum* SnAGAT1 (Lou et al., 2021). Both enzymes reside outside of the ASAT3 clade, and thus are phylogenetically distinct from SmASAT3-L1. SmASAT3-L1 characteristics and those of recently published *S. quitoense* enzymes (Leong et al., 2022, 2020) also suggest that acylinositols are synthesized through one or more pathway(s) distinct from that of acylsucroses. Existing genetic resources in *S. nigrum*, *S. quitoense*, and *S. melongena* can be utilized to explore when and how acylinositols arose and to identify the underlying mechanisms behind acylation position and chain type differences between these species.

The remarkably varied structures of acylsugars strongly suggest that they have evolved distinct bioactivities, likely in coevolutionary arms races between plants and their insect and microbial pests. For example, hydrophobic acylsugars with longer acyl chains and few free hydroxyl groups may disrupt membranes, reminiscent of less polar triterpene saponin variants (Augustin et al., 2011; Baumann et al., 2000). Variation in acyl chain linkage chemistries also likely influence acylsugar mode of action. For example, ester linked short chains acylsucroses are digested by

Manduca sexta (hawkmoth) larvae and the volatile organic acids released into the environment, attracting ant predators (Weinhold and Baldwin, 2011). In contrast, longer chain hydroxylated and ether-linked glycohydroxyacyl chains observed in the study are likely to have quite different metabolic fates, perhaps persisting in the digestive systems of herbivores, and alternative metabolic fates in microbes. Uncovering the enzymes responsible for synthesis of the wide variety of natural acylsugars will enable transgenic production of isogenic variants and rigorous testing of their ecological activities. Similarly, access to a wider variety of often promiscuous ASAT BAHD acyltransferases (Moghe et al., 2023) creates opportunities for producing novel acylsugars with a wide range of physical properties and bioactivities (Schenck et al., 2022).

Hydroxyacyl chains also present promising opportunities to expand upon natural acylsugar diversity through synthetic biology and/or synthetic chemistry. The hydroxyl acts as a reactive chemical handle, which was exploited in a subset of *Solanum* species trichomes to create glycohydroxyacylhexoses. This handle can also be exploited with synthetic chemistry using these chains as new feedstocks for creating polymers. These chains may also be useful in creating synthetic acylsugars for use as food-grade emulsifiers. We can also further modify the hydroxyacyl chains to add unusual sugars, acyl chains, and aromatic groups through promiscuous enzymes or synthetic chemistry. These structural changes would likely impact biological activities and add another asset to developing more pest resistant plants. Characterizing more natural acylsugar diversity will further enable synthetic biology approaches to create completely new acylsugar structures.

Methods

Plant material and growth conditions

Solanum spp. seeds were obtained from the sources described in Table S37. Seeds were treated with 10% (v/v) bleach for 10 min while being rocked at 24 rpm with a GyroMini nutating mixer (Labnet, Edison, NJ, USA), and then rinsed 5-6 times with distilled water. Unless otherwise noted, seeds were germinated on Whatman filter paper (MilliporeSigma, Burlington, MA, USA) at 28°C and in the dark. Germinated seedlings were transferred to peat pots (Jiffy, Zwijndrecht, Netherlands), and grown at 25°C, 16/8-h day/night light cycle, and ~70 µmol $m^{-2}s^{-2}$ photosynthetic photon flux density with cool white fluorescent bulbs. Mature plants were grown in controlled environment growth chambers or in a greenhouse. The growth chamber conditions consisted of 25°/18°C day/night temperatures, 16/8-h day/night light cycle, and ~100 µmol $m^{-2}s^{-2}$ photosynthetic photon flux density under LED bulbs. The greenhouse conditions consisted of a 16/8-h day/night light cycle achieved with supplemental sodium iodide lighting. Plants were fertilized weekly with 0.5X Hoagland's solution.

Surface metabolite extractions

Surface metabolites were extracted as described previously (Leong et al., 2019; Lou and Leong, 2019). Briefly, 0.1-1 g of leaf, stem, hypocotyl, or cotyledon tissue was collected into a 1.5 mL screw-cap tube (Dot Scientific, Inc., Burton, MI, USA), 1 mL extraction solvent (3:3:2 acetonitrile:isopropanol:water, 0.1% formic acid, 10 µM propyl 4-hydroxybenzoate (internal standard)) was added and the sample was rocked at 24 rpm by a GyroMini nutating mixer for two min. After extraction, the supernatant was transferred into a glass 2 mL autosampler vial (Restek, Bellefonte, PA, USA) and sealed with a 9 mm cap containing a PTFE/silicone septum (J.G. Finneran, Vineland, NJ, USA). For fruit surface metabolite extractions, we modified the

protocol by placing 1-2 fruit into a 15 mL polypropylene conical tube (Corning Inc., Corning, NY, USA) containing 5 mL extraction solvent then proceeded to the nutation step above.

Bulk eggplant acylsugar extraction and purification for NMR analysis

Surface metabolites were bulk extracted from approximately 1000 *S. melongena* PI 555598 seedlings. Seeds were treated with 10% (v/v) bleach (Clorox, Oakland, CA, USA) for 10 min while being gently rocked at 24 rpm with a GyroMini nutating mixer and subsequently rinsed 5-6 times with distilled water. Seeds were sown in moist SUREMIX soil (Michigan Grower Products, Galesburg, MI, USA) in Perma-Next plant trays, 22 x 11 x 2.5 inches (Growers Supply Company, Dexter, MI, USA), covered with a humidity dome, 22 x 11 x 3 inches (Growers Supply Company), then transferred to the growth chamber immediately. Seedlings were harvested when 2-3 true leaves were observed, approximately one week, by cutting them at the base and placing them into two 2L beakers each containing 1L 100% acetonitrile. Surface metabolites were extracted with gentle agitation with a metal spatula for five min at room temperature. Plant material and sediment was removed by vacuum filtration through a Büchner funnel lined with Whatman filter paper (MilliporeSigma). solvent was removed *in vacuo* by rotary evaporation and dried residue was dissolved in 20 mL of acetonitrile with 0.1% formic acid. Solvent was removed again using a vacuum centrifuge and the dried residue was dissolved in 1 mL of acetonitrile:water:formic acid (80:20:0.001). The semi-preparative LC method is described in detail in Table S32 and is described in brief here. *S. melongena* acylsugars were separated with a Waters 2795 HPLC (Waters Corporation, Milford, MA, USA) equipped with an Acclaim 120 C18 HPLC column (4.6 x 150 mm, 5 µm; Thermo Fisher Scientific, Waltham, MA, USA). Solvent A was water with 0.1% formic acid and Solvent B was acetonitrile. The reverse-

phase linear gradient was as follows: 5% B at 0 min, 60% B at 2 min, 80% B at 40 min, 100% B at 42 min, 5% B at 42.01 min, held at 5% B until 44 min. Flow rate was 1.5 mL/min, injection volume was 100 µL, and the column temperature was 40°C. Fractions were collected automatically at 0.25 min intervals by a 2211 Superrac fraction collector (LKB Bromma, Stockholm, Sweden) and assessed for acylsugar presence and purity by LC-MS analysis, as described below. Column fractions were collected in the same tubes for each method run which worked to pool the fractions between each run.

LC-MS acylsugar analysis

Acylsugars were analyzed by LC-MS with each method described below and in Tables S27-31. For each LC method, the mobile phase consisted of aqueous 10 mM ammonium formate, adjusted to pH 2.8 with formic acid (Solvent A) and 100% acetonitrile (Solvent B). The flow rate was maintained at 0.3 mL/min.

Acylsugar extracts were analyzed with a 22 min LC gradient using a Waters Acquity UPLC coupled to a Waters Xevo G2-XS QToF mass spectrometer (Waters Corporation, Milford, MA) equipped with electrospray ionization (ESI) operating in positive (ESI^+) and negative (ESI^-) modes. 10 µL acylsugar extracts were separated on an Acquity UPLC BEH C18 column (10 cm x 2.1 mm, 130 Å, 1.7 µm; Waters), kept at 40°C, using a binary solvent gradient. The 22 min linear gradient was as follows: 5% B at 0 min, 60% B at 2 min, 100% B at 16 min, held at 100% B until 20 min, 5% B at 20.01 min, held at 5% B until 22 min. Acylsugar extracts were analyzed with both ESI^-, and ESI^+. For ESI^-, the following parameters were used: capillary voltage, 2 kV; sampling cone voltage, 35 V; source temperature, 100°C; desolvation temperature 350°C; cone

gas flow, 50 L/Hr; desolvation gas flow, 600 L/Hr. For ESI^+, the following parameters were used: capillary voltage, 3 kV; sampling cone voltage, 35 V; source temperature, 100°C; desolvation temperature 300°C; cone gas flow, 50 L/Hr; desolvation gas flow, 600 L/Hr. Acylsugars were fragmented in either MS^E or data-dependent acquisition (DDA) MS/MS modes as described previously (Lou et al., 2021; Lybrand et al., 2020). DDA survey and MS/MS functions acquired over *m/z* 50 to 1500 with scan times of 0.2 s. Ions selected for MS/MS were fragmented with a ramped collision energy where voltage varies based on ion mass. Collision energies followed a ramp with the voltage changing linearly for ions between the low mass setting (*m/z* 50, 15 to 30 V) and high mass setting (*m/z* 1500, 30 to 60 V). To increase mass accuracy, lock mass correction was performed during data collection, with leucine enkephalin as the reference.

Semi-preparative LC fractions were analyzed with direct infusion and a 14-min LC gradient using a LC-20ADvp ternary pump (Shimadzu, Kyoto, Japan) coupled to a Waters Xevo G2-XS QToF mass spectrometer (Waters Corporation). The direct infusion method quickly screened fractions for acylsugar presence while the 14-min LC method tested fraction purity. 10 µL of acylsugar fractions were injected into an Ascentis Express C18 HPLC column (10 cm x 2.1 mm, 2.7 µm; Supelco, Bellefonte, PA, USA), kept at 40°C. The 14-min gradient was as follows: 5% B at 0 min, 60% B at 2 min, 100% B at 10 min held until 12 min, 5% B at 12.01 min, and held at 5% until 14 min. Fractions were analyzed under ESI^+ with the following parameters: capillary voltage, 3 kV; sampling cone voltage, 40 V; source temperature, 100°C; desolvation temperature 350°C; cone gas flow, 20 L/hr; desolvation gas flow, 500 L/hr. For direct infusion analysis, ions were acquired from *m/z* 50 to 1500 with a scan time of 0.1 s and one acquisition function with 0

V collision potential. For 14-min LC analysis, ions were acquired from *m/z* 50 to 1200 with a scan time of 0.1 s and three acquisition functions with different collision potentials (0, 25, 60 V). Lock mass calibration referenced to the leucine enkephalin $[M+H]^+$ ion was applied during data acquisition.

Enzyme assays products were analyzed with a 7-min gradient using Waters Acquity UPLC coupled to a Waters Xevo G2-XS QToF mass spectrometer (Waters Corporation, Milford, MA) equipped with electrospray ionization (ESI) operating in negative (ESI^-) mode. Reaction products were separated with an Ascentis Express C18 HPLC column (10 cm x 2.1 mm, 2.7 µm; Supelco), kept at 40°C, using a binary solvent gradient. The 7 min linear gradient was as follows: 5% B at 0 min, 60% B at 1 min, 100% B at 5 min, held at 100% B until 6 min, 5% B at 6.01 min, held at 5% B until 7 min. The ESI^- parameters described for acylsugar extract analysis were used. Ions were acquired from *m/z* 50 to 1200 with a scan time of 0.1 s and three acquisition functions with different collision potentials (0, 25, 60 V). Lock mass calibration referenced to the leucine enkephalin $[M+H]^-$ ion was applied during data acquisition.

For coelution analysis between enzymatically- and plant-produced AI3:16 and AI4:18, a 24-min linear gradient was used with an Acquity UPLC BEH C18 column (10 cm x 2.1 mm, 130 Å, 1.7 µm; Waters), kept at 40ºC, on the same instrument used for enzyme assay analysis. The binary solvent, linear gradient was as follows: 5% B at 0 min, 60% B at 18 min, 100% B at 18.01 min, held at 100% B until 22 min, 5% B at 22.01 min, held at 5% B until 24 min. ESI^- parameters and MS method from the above enzyme assay analysis were used.

Saponified acylsugars were analyzed using a Waters Acquity UPLC coupled to a Waters Xevo G2-XS QToF mass spectrometer (Waters Corporation, Milford, MA). 10 µL of saponified acylsugars were injected into either an Acquity UPLC BEH C18 column (10 cm x 2.1 mm, 130 Å, 1.7 µm; Waters) or an Acquity UPLC BEH Amide column (10 cm x 2.1 mm, 130 Å, 1.7 µm; Waters), kept at 40°C. Samples were analyzed on the C18 column with the 22-min method described above. To detect free sugars resulting from saponification, samples were analyzed on the BEH amide column with a 9 min method which was as follows: 95% B at 0 min held until 1 min, 60% B at 6 min, 5% B at 7 min, 95% B at 7.01 min, held at 95% B until 9 min. Both methods used the ESI- parameters and MS acquisition parameters described for enzyme assays.

Acylsugar annotation

We inferred acylsugar structures utilizing negative and positive mode MS and MS/MS collision-induced dissociation (CID) as previously described (Ghosh et al., 2014a; Hurney, 2018; Leong et al., 2019; Lou et al., 2021; Lybrand et al., 2020). We employed modified confidence criteria developed by Lou and coworkers (Lou et al., 2021). Putative acylsugars only meeting criterion A were annotated as low confidence. Putative acylsugars meeting criteria A-D were annotated as medium confidence. We also added two additional criteria, G and H, described below. Acylsugars meeting criteria A-F and either G or H were annotated with highest confidence. Confidence levels for each reported acylsugar are specified in their species-specific annotation table (Tables S1-26).

(G) **NMR structural determination.** Acylsugars from *S. melongena* were further characterized by NMR experiments in which sugar core relative stereochemistry, acyl chain positions, and acyl chain branching were resolved.

(H) Coelution with NMR characterized compounds. Acylsugars with the same acyl chains and sugar core size were analyzed by coelution with *S. melongena* acylsugars. If acylsugars coeluted with an NMR characterized peak and shared the same acyl chains and fragmentation patterns, they met this criterion and were annotated with high confidence. This criterion was employed for acylsugars from species other than *S. melongena*.

Acylsugar metabolomic analysis

To collect accurate formate adduct masses and relative abundances from putative acylsugars, negative mode LC-MS raw data from metabolite extracts were analyzed with Progenesis QI software v3.0 (Waters, Milford, MA, USA) using retention time alignment, peak detection, adduct grouping, and deconvolution. Lock mass correction was performed during data collection. We used the following analysis parameters: peak picking sensitivity, default; retention time range, 0-22 min; adducts detected, M+FA-H, M+Cl, M-H, 2M-H, and 2M+FA-H. Average percent peak abundance was calculated for each putative acylsugar by dividing its raw abundance by the sum of the total acylsugar raw abundance and then averaging that value across all samples for a species.

NMR analysis of acylsugars

Purified acylsugars were dissolved in acetonitrile-d_3 (99.96 atom % D; MilliporeSigma, Burlington, MA, USA) and transferred to Kontes NMR tubes (MilliporeSigma, Burlington, MA, USA). Samples were analyzed at the Michigan State University Max T. Rogers NMR Core with a Varian Inova 600 MHz spectrometer (Agilent , Santa Clara, CA, USA) equipped with a Nalorac 5 mm PFG switchable probe, a DirectDrive2 500 MHz spectrometer (Agilent, Santa Clara, CA, USA) equipped with a OneNMR probe, Bruker Avance NEO 800 MHz spectrometer (Bruker, Billerica, MA, USA) equipped with a 5 mm helium cryogenic HCN probe, or a Bruker Avance NEO 600 MHz spectrometer equipped with a 5 mm nitrogen cryogenic HCN Prodigy probe. All acylsugars were analyzed with a series of 1D (^{1}H, and ^{13}C) and 2D (gCOSY, TOCSY, gHSQC, gHMBC, gH2BC, *J*-resolved) experiments. Resulting spectra were referenced to acetonitrile-d_3 ($^{\delta}$H = 1.94 and $^{\delta}$C = 118.70 ppm). NMR spectra were processed and analyzed with MestReNova 12.0 software (MestreLab, A Coruña, Spain). For full NMR metadata, see Supplementary Data.

Sugar core composition analysis

S. melongena PI 555598 acylsugar extracts were dried down *in vacuo* with centrifugation (Savant, ThermoFisher Scientific). Dried acylsugar extracts were dissolved in 1 mL of methanol. One mL of 3 M ammonium hydroxide was added, and the solution was mixed vigorously. This initial saponification reaction proceeded at room temperature for 48 hours. Solvent was removed by vacuum centrifugation. Saponified sugar cores were reduced and derivatized with acetate groups as previously described (Sassaki et al., 2008). Sugars were dissolved in 50 µL of 1 M ammonium hydroxide for 15 min at room temperature. Addition of 50 µL 20 mg/mL $NaBH_4$ and incubation at 100°C for 10 min converted aldoses and ketoses to polyols. Excess sodium

borohydride was quenched with 100 µL 1M trifluoroacetic acid. Two volumes of methanol were added, and the sample was dried down by a stream of nitrogen gas. Dried down reaction products were redissolved in 200 µL of methanolic 0.5 M HCl and heated at 100°C for 15 min in a closed ½ dram vial, 12 x 35 mm (Kimble Chase, Vineland, NJ, USA). Solvent was removed by a stream of nitrogen gas. To acetylate the sugar cores, residue was dissolved in 200 µL of pyridine:acetic anhydride (1:1) and incubated at 100°C for 30 min. Acetylated polyols were then dried down *in vacuo* in a vacuum centrifuge and redissolved in 100 µL hexane for GC-MS analysis.

To determine the composition of disaccharide sugar cores, saponified sugars were hydrolyzed prior to derivatization. Dried down sugars were redissolved in 98 µL water in a ½ dram vial, 12 x 35 mm (Kimble Chase). After addition of 102 µL of 88% formic acid, the vial was sealed and heated at 100°C for 15 hours in a heat block. Hydrolyzed sugars were dried down *in vacuo* with a vacuum centrifuge and derivatized to alditol acetates following the method described above.

Hydroxyl acyl chain stereochemistry analysis

We derivatized and prepared hydroxyacyl chains as previously described (Jenske and Vetter, 2007). The commercial standards (3*R*)-OH-tetradecanoate (Cayman Chemical, Ann Arbor, MI, USA) and (3*R*/*S*)-OH-tetradecanoate (Cayman Chemical, Ann Arbor, MI, USA) were first derivatized to their respective ethyl ester derivatives. 600 µg of each fatty acid was dissolved in 0.5 mL of 0.5 M ethanolic KOH and incubated at 80°C for 5 min. After the solution was cooled on ice, 1 mL of ethanolic BF_3 was added, and reactions were incubated at 80°C for five min. Two-phase partitioning with hexane and saturated sodium chloride yielded ethyl ester derivatives in the hexane partition.

The 3-OH-nC14 acyl chain from I3:22(iC4,iC4,3-OH-nC14) was derivatized to an ethyl ester following a previously published method (Fan et al., 2020; Ning et al., 2015). Half of the purified I3:22(iC4,iC4,3-OH-nC14) was dissolved in acetonitrile with 0.1% formic acid and transferred to a 1.5 mL microcentrifuge tube. The purified acylsugar was dried down *in vacuo* with vacuum centrifugation. The resulting dry, purified I3:22(iC4,iC4,3-OH-nC14) was dissolved in 300 µL of 21% (w/w) sodium ethoxide in ethanol (MilliporeSigma, Burlington, MA, USA). The reaction proceeded at room temperature for 30 min with constant rocking at 24 rpm by a GyroMini nutating mixer (Labnet, Edison, NJ, USA) and occasional vortexing. 400 µL hexane with 55 ng/µL tetradacane (MilliporeSigma, Burlington, MA, USA) as an internal standard was then added followed by vigorous vortexing. 500 µL of aqueous saturated sodium chloride was added to the hexane-ethanol mixture, and vortexed. The hexane layer was pipetted to another tube and two more two-phase partitions with saturated sodium chloride were completed. The final hexane layer containing the acyl chain ethyl esters was extracted and dried down by a stream of nitrogen gas.

The resulting fatty acid ethyl esters from the commercial standards and purified compound were derivatized with (*R*)-(-)-α-methoxy-α-trifluoromethylphenylacetyl chloride ((*R*)-(-)-MTPA-Cl; MilliporeSigma, Burlington, MA, USA) following a previously published procedure (Jenske and Vetter, 2007). Dried fatty acid ethyl esters were redissolved in 400 µL pyridine and 15 µL of (*R*)-(-)-MPTA-Cl. The reaction proceeded at room temperature for two hours, after which, 5 mL of water and tert-butyl methyl ether (TBME) were added along with solid K_2CO_3 (one spatula

tip). The TBME phase was collected after three successive phase separations with 5 mL of water. TBME was evaporated to 1 mL and subjected to GC-MS analysis.

GC-MS analysis

All GC-MS analyses employed an Agilent 5890 GC and an Agilent 5975 single quadrupole MS equipped with a FactorFour VF-5ms column (30 m x 0.25 mm, 0.25 um; Agilent) and 10 m EZ-Guard column (Agilent, Santa Clara, CA, USA). Helium was used as the carrier gas with a constant flow rate of 1 mL/min. Electron energy was set at 70 eV. MS source and quadrupole were maintained at 230°C and 150°C, respectively. Parameters specific to each analysis are listed below.

For analysis of sugar core alditol acetate derivatives, we followed previously published GC-MS parameters (Sassaki et al., 2008). Inlet temperature was maintained at 275°C. GC oven temperature was held at 60°C for one min and then was ramped at a rate of 40°C/min to 180°C. Oven temperature was then ramped to 240°C at a rate of 5°C/min and held for three min. Total run time was 19 min. The MS detector transfer line was maintained at 280°C. Split ratio was 10:1 with a split flow rate of 10 mL/min. A three min solvent delay was used. MS data was collected in full scan mode, *m/z* 50-600.

For analysis of MPTA fatty acid ethyl ester derivatives, we followed previously published GC-MS parameters (Jenske and Vetter, 2007). GC oven temperature was held at 60°C for 1.5 min and then was ramped at a rate of 40°C/min to 180°C. 180°C was held for two min after which the oven temperature was ramped to 230°C at a rate of 2°C/min. After 230°C was held for nine

min, the oven temperature was ramped to 300°C at 10°C/min and held for 7.5 min. Total run time was 55 min. A 5-min solvent delay was applied. The MS detector transfer line was maintained at 280°C. Selective ion monitoring detected the ions *m*/*z* 189, 209, and 255 with a dwell time of 50 ms.

Sample collection and transcriptome sequencing

We sequenced 18 transcriptomes: six biological replicates each of trichomes isolated from hypocotyls (trichomes), hypocotyls stripped of their trichomes (stripped hypocotyls), and whole roots (Supplemental data file 1). We sampled trichomes and stripped hypocotyls from 7-day-old *S. melongena* 67/3 seedlings following methods developed for root hair cell isolation (Bucher et al., 1997) with modifications. We grew lawns of eggplant seedlings in soil flats as described above for bulk leaf surface metabolite extraction. At 7 days post germination, we removed roots and cotyledons from seedlings, then transferred hypocotyls to liquid nitrogen in a plastic 2 L Dewar flask. Frozen hypocotyls were gently stirred with a glass rod for 20 minutes to physically shear trichomes from hypocotyls. After confirming under a dissecting scope that a sample of three hypocotyls had been stripped bare, we filtered trichomes into a 2 L glass beaker by slowly pouring the contents of the Dewar flask through a 500 µm wire mesh sieve (MilliporeSigma, Burlington, MA, USA). To maximize trichome recovery, stripped hypocotyls were returned to the Dewar, rinsed with liquid nitrogen, and filtered through the 500 µm sieve six more times. Stripped hypocotyls were divided into six pre-weighed 50 mL conical tubes, quickly weighed, then transferred to -80ºC for storage. Trichomes were subsequently filtered through a 150 µm sieve (MilliporeSigma, Burlington, MA, USA) into a 500 mL beaker to increase sample purity then transferred to a 50 mL conical tube to allow the excess liquid nitrogen to evaporate. Finally,

filtered trichomes were divided into six pre-weighed 2 mL screw-cap tubes (Dot Scientific, Inc., Burton, MI, USA), quickly weighed, then transferred to -80ºC for storage.

We extracted total RNA using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's directions, measured RNA concentration using a Qubit 1.0 instrument (Thermo Fisher Scientific, Waltham, MA, USA) with the RNA HS assay, and the samples were processed by the Michigan State University Research Technology Support Facility Genomics Core (East Lansing, MI, USA) for library preparation and high-throughput sequencing. The core checked RNA quality using a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA), constructed sequencing libraries using an Illumina Stranded mRNA Prep kit (Illumina, Inc. San Diego, CA, USA), and sequenced all 18 libraries on a NovaSeq 6000 (Illumina, Inc. San Diego, CA) using a single S4 flow cell lane, producing 150-base pair paired-end reads (Supplemental datafile 1). The raw fastq files are available to download from the National Center for Biotechnology Information Sequence Read Archive database under BioProject PRJNA935765.

Transcriptome alignment and differential gene expression analysis

To prepare the 150-bp paired-end sequences for alignment, we employed Trimmomatic (Bolger et al., 2014) to trim adapters and low-quality bases, then filter reads shorter than 75 bp, removing an average of 3.7% of sequences (range: 3.3% - 4.2%). We mapped the resulting 75–150-bp (average length: 149 bp) paired-end RNAseq reads separately to the eggplant V4.1 and HQ reference genomes using STAR in two-pass mode, which enhances splice junction discovery and mapping sensitivity (Dobin et al., 2013; Dobin and Gingeras, 2015). Using this approach, an average of 79.2% (range: 60.1 - 87.1%) and 82.2% (range: 62.0 - 90.7%) of reads mapped to a

unique genomic location in V4.1 and HQ, respectively (Table S38). We filtered the resulting transcriptome alignments according to best practices as defined by the Genome Analysis Toolkit (GATK) (DePristo et al., 2011; Van der Auwera et al., 2013). Briefly, we removed optical and PCR duplicates with MarkDuplicates from the Picard toolkit (http://broadinstitute.github.io/picard), parsed reads into exon segments and removed intron-spanning bases using SplitNCigarReads from GATK (McKenna et al., 2010). Finally, we selected unique alignments by eliminating reads with a mapping quality score below Q60 with the view command from SAMtools (Li et al., 2009). To generate raw read counts, we used the HTSeq command, htseq-count with the nonunique parameter set to all (Anders et al., 2015).

Raw read counts generated by HTSeq-count were used to perform differential gene expression analysis in edgeR (Robinson et al., 2010). To restrict comparisons to expressed genes, only transcripts with at least one read count-per-million (CPM) in at least one sample were retained for further analysis. This filtering step removed 11,665 (33.4%) and 14,544 (39.8%) annotated genes in the V4.1 and HQ eggplant genomes, respectively, yielding 23,251 and 22,024 expressed transcripts for differential gene expression analysis. Next, we normalized transcript abundances across the 18 transcriptomes in our eggplant V4.1 and HQ alignments using the default trimmed mean of M-values (TMM) method in the calcNormFactors function, then performed multidimensional scaling (MDS) with the plotMDS function to compare global gene expression profiles (Figure S13). This showed that our samples cluster tightly by tissue (i.e., trichome, shaved hypocotyl, or root), with no obvious differences between V4.1 and HQ alignments. To test differences in gene expression across tissues, we implemented a generalized linear model (GLM) using a quasi-likelihood (QL) approach: we generated an experimental design matrix

specifying the three tissues (i.e., trichomes, trichomeless hypocotyls, and roots) with the model.matrix function, then used the glmQLFit function to fit our data to a QL-GLM model. To identify genes with a log2 fold-change (FC) > 2 between tissues, we used the glmTreat function, which performs threshold hypothesis testing, a rigorous statistical approach that evaluates variance and magnitude to detect expression differences greater than the specified value (e.g., log2 FC > 2), then applies false discovery rate (FDR) *p*-value corrections. Genes were classified as significantly differentially expressed between two tissues if log2 FC > 2 and FDR-corrected *p*-value < 0.05. We calculated absolute transcript abundance for all expressed genes as transcripts per million (TPM) with the calculateTPM function in scateR (McCarthy et al., 2017). Transcriptome expression values are in Tables S39 and S40.

Phylogenetic analyses

Characterized ASAT enzymes fall into clade III of the BAHD family (Fan et al., 2017, 2015; Leong et al., 2022, 2020; Lou et al., 2021; Moghe et al., 2017; Nadakuduti et al., 2017; Schenck et al., 2022; Anthony L Schilmiller et al., 2015; Schilmiller et al., 2012). To identify ASAT candidates in eggplant, we searched for sequences containing the Pfam Transferase domain (PF02458) (Finn et al., 2010), associated with all characterized catalytically active BAHD proteins. The PF02458 HMM profile was obtained from the Pfam website (http://pfam.sanger.ac.uk/) and queried against the V4.1 and HQ eggplant proteomes using the hmmsearch tool from HMMER v3.2.1 (hmmer.org), revealing 106 and 108 putative BAHD sequences, respectively. Using MAFFT v7.471 in E-INS-i mode, we built multiple sequence alignments (MSA) of amino acid sequences from four sources: 1) V4.1 or HQ eggplant PF02458 hits, 2) published reference sequences for clades I-VII (Moghe et al., 2023), 3) characterized

ASAT sequences from other Solanaceae species, and 4) the SaASAT3 and SaASAT3-L1 candidates from *S. aethiopicum* (Katoh and Standley, 2013). The E-INS-i algorithm implements local alignment with a generalized affine gap cost (Altschul, 1998), which aligns conserved regions (e.g., the BAHD transferase domain) and essentially ignores nonconserved regions. Phylogenetic reconstruction was performed using IQ-TREE v2.1.3 (Minh et al., 2020). The ModelFinder tool was implemented to identify the best maximum likelihood model for estimating evolutionary relationships (Kalyaanamoorthy et al., 2017), leading to selection of Jones-Taylor-Thornton (JTT)+F with seven rate categories and Q.plant+F with seven rate categories for the V4.1 and HQ MSAs, respectively. Phylogenetic trees were inferred by maximum likelihood using the chosen model, and branch support was obtained from 100,000 ultrafast bootstrap iterations (Hoang et al., 2018). The resulting BAHD phylogenies were visualized using the ggtree package in R (Yu et al., 2017). To generate the clade III BAHD heatmap-tree, we used the viewClade function in ggtree to subset the phylogeny and used gheatmap to visualize transcript abundance (log2 TPM) for the eggplant BAHDs.

BAHD acyltransferase cloning, expression, and purification

Candidate ASATs were cloned into pET28b(+) (MilliporeSigma, Burlington, MA, USA). Open reading frames from genes were either synthesized by Twist Biosciences (South San Francisco, CA, USA) with or without codon-optimization for *E. coli* expression (Table S35) or amplified from genomic DNA or cDNA with primers listed in Table S36. Q5 2X Hotstart master mix (New England Biolabs, Ipswich, MA, USA) was used for cloning PCRs. The source of cloned DNA and whether a gene was codon optimized is described in Table S35. Amplified genes were purified by agarose gel electrophoresis and extraction with the Monarch DNA Gel Extraction Kit

(New England Biolabs, Ipswich, MA, USA). Both the synthesized genes and PCR amplified genes were then inserted into a doubly digested BamHI/XhoI pET28b(+) through Gibson assembly using the 2X Gibson Assembly Master Mix (New England Biolabs, Ipswich, MA, USA) according to the manufacturer's instructions. The constructs using synthesized genes were transformed into BL21(DE3) (MilliporeSigma, Burlington, MA, USA) and constructs with PCR amplified genes were transformed into BL21 Rosetta(DE3) cells (MilliporeSigma, Burlington, MA, USA). Constructs were verified with colony PCR and Sanger sequencing using T7 terminator and promoter primers (Table S36). Sanger sequencing was completed by the Michigan State University Research Technology Support Facility Genomics Core (East Lansing, MI, USA).

Protein expression occurred as previously described (Leong et al., 2022; Lou et al., 2021). Briefly, 50 mL cultures of picked transformation colonies were grown overnight at 37°C, shaking at 225 rpm in Luria-Bertani (LB) media (Neogen, Lansing, MI, USA) supplemented with 1% glucose (w/v). Fifteen mL of the overnight cultures were inoculated into 1 L of fresh LB medium, which was incubated at 37°C shaking at 225 rpm until an OD600 of 0.5 was reached. The cultures were incubated on ice for 25 min, after which, isopropylthio-β-galactoside was added to a final concentration of 50-500 µM. Then, cultures were incubated at 16°C shaking at 180 rpm overnight for 16 hours before cells were harvested by centrifugation at 4,000 rpm for 10 minutes at 4°C.

S. melongena BAHDs were purified as previously described (Leong et al., 2020) with the following modifications. The extraction buffer contained 10 mM imidazole, the wash buffer

contained 20 mM imidazole, and the elution buffer contained 500 mM imidazole. Protein eluent was concentrated with 30-kD Amicon Ultra centrifugal filter units (MilliporeSigma).

Enzyme assays

Enzyme assays were conducted in 100 mM sodium phosphate buffer at pH 6. For forward assays, acetyl-CoA (MilliporeSigma, Burlington, MA, USA) was added to a final concentration of 0.1 mM, and for reverse assays free CoA (MilliporeSigma, Burlington, MA, USA) was added to a final concentration of 1 mM. Purified AI3:16 and AI4:18 substrates were dried down using a vacuum centrifuge and redissolved in ethanol:water:formic acid (1:1:0.001). One microliter of the prepared acylsugars were used as acyl acceptors. Six microliters of enzyme were added to a final volume of 60 µL. For negative controls, 6 µL of enzyme that was heat inactivated at 95°C for 10 minutes was substituted in place of untreated enzyme. Assays were incubated at 30°C for 30 minutes after which 120 µL of acetonitrile:isopropanol:formic acid (1:1:0.001) with 1.5 µM telmisartan (MilliporeSigma, Burlington, MA, USA) stop solution was added. Reactions were then spun at 17,000 x *g* for 10 minutes to remove precipitate. Supernatant was placed in autosampler vials and analyzed by LC-MS.

RT-PCR of *S. aethiopicum* BAHDs

We employed semi-quantitative RT-PCR to test *S. aethiopicum* BAHD expression in glandular trichomes with cDNA and genomic DNA (gDNA). Total RNA was isolated with the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany), including an on-column DNase digestion (Qiagen, Hilden, Germany), from shaved hypocotyls from accession *S. aethiopicum* PI 666075 and glandular trichomes from accessions PI 666075 and Grif 14165 isolated as described above for *S.*

melongena glandular trichomes. RNA was quantified with a Nanodrop 2000c instrument (Thermo Fisher Scientific, Waltham, MA, USA). We synthesized cDNA using 10 ng of RNA and SuperScript III Reverse Transcriptase (Invitrogen, Waltham, MA, USA). gDNA was isolated with the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) from young leaf tissue collected from mature *S. aethiopicum* PI 666075 plants. PCR reactions (25 µL) were set up with GoTaq Green Master Mix (Promega, Madison, WI, USA), 200 nM of forward and reverse primers (Table S36), and 1 µL of cDNA or gDNA. PCR was performed under these conditions: 2 min at 95°C followed by 22, 30, and 35 cycles of 30s at 95°C, 30s at 58°C, and 1 min at 72°C.

We identified the putative *S. aethiopicum* orthologs of the *S. melongena* candidate BAHDs by querying *S. aethiopicum* annotated transcripts (Song et al., 2019) with *S. melongena* candidate ASAT DNA sequences using BLASTn. SaASAT3 and SaASAT3-L1 (GenBank accession: OQ547782) were not annotated in the *S. aethiopicum* genome and were identified by querying *S. aethiopicum* scaffolds with BLASTn and then identifying open reading frames with Geneious software v9.1.8 (Dotmatics, Boston, MA, USA). The *S. aethiopicum* ef1α gene was identified by querying a putative *S. melongena* ef1α, SMEL4.1_06g005890, against the *S. aethiopicum* scaffolds.

<u>*S. melongena* acyl chain composition analysis</u>

S. melongena acyl chain composition was assessed with a previously developed esterification and GC-MS methods (Fan et al., 2020; Ning et al., 2015; Schenck et al., 2022). Acyl chain identifies were determined through authentic reference standards for nC8, nC10, nC12, and nC14 fatty acid ethyl esters (MilliporeSigma). The QuanLynx function of MassLynx v4.1 (Waters

Corporation) integrated peaks from extracted ion chromatograms. Extracted ion chromatograms for the medium acyl chains iC8, nC8, iC10, nC10, iC12, nC12, iC14, and nC14 were generated for the *m/z* of 88 with a mass window of *m/z* of 0.50. Extracted ion chromatograms for the short chains iC4, aiC5, and iC5 were generated for with the *m/*z values of 71, 102, and 101, respectively, with a mass window of *m/z* of 0.50.

Accession numbers

Raw RNA sequence data can be found in the National Center for Biotechnology Information (NCBI) Sequence Read Archive database under BioProject PRJNA935765. Sequence data for SaASAT3-L1 is available from the NCBI GenBank database under accession number OQ547782.

Acknowledgement

This research was funded by the National Science Foundation grant IOS-PGRP-1546617 (to R. L. L. and A. D. J.), National Science Foundation grant IOS-PGRP-2218206 (to R. L. L.) and National Institute of General Medical Sciences of the National Institutes of Health graduate training grant no. T32-GM110523 (to P. D. F.). A. D. J. is supported in part by USDA National Institute of Food and Agriculture Hatch project number MICL02474. We thank Dr. Christopher T. Martine for providing Australian *Solanum* ssp. seeds, Dr. Joyce van Eck for providing *Solanum prinophyllum* seeds, Dr. Guiseppe Leonardo Rotino and the Consiglio per la ricercar in agricoltura e l'analisi dell'economia agraria for providing *S. melongena* 67/3 and 305E40 seeds, and USDA-GRIN for providing *Solanum* ssp. seeds. We acknowledge the Michigan State University RTSF Mass Spectrometry and Metabolomics Core Facilities for LC-MS analysis support. We also acknowledge Dr. Daniel Holmes and Dr. Li Xie at the Michigan State University Max T. Rogers NMR Facility for experimental design and data analysis support. We also acknowledge members of the Last lab for helpful feedback.